Estimating the association between air pollution exposure and mortality using Bayesian hierarchical models

Lee, Duncan Paul

Award date:
2007

Awarding institution:
University of Bath

Estimating the association between air pollution exposure and mortality using Bayesian hierarchical models

submitted by

Duncan Paul Lee

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Mathematical Sciences

January 2007

Summary

This thesis develops statistical methodology for an important area of environmental epidemiology, that of the relationship between short-term exposure to air pollution and mortality or morbidity which has been a public health concern for over fifty years. The majority of studies investigating this relationship are based on ecological data, and estimate a group level association between ambient pollution levels and population aggregated mortality. This association is typically estimated with Poisson regression models, which make a number of simplifying assumptions about the underlying processes that generate the data. The work presented in this thesis extends the standard approaches to modelling these data in three main ways, the first proposing the use of autoregressive processes rather than smooth functions to remove any long-term trends and temporal correlation in the daily mortality series. The second extension relates to the pollution-mortality relationship, and investigates whether it changes over time rather than being constant or a dose-response curve. The remainder of this thesis investigates the importance of correctly estimating pollution exposures, and how mis-estimating them affects the resulting health risk. These extensions are implemented using Bayesian hierarchical models with estimation achieved via Markov chain monte carlo simulation. For the first two extensions likelihood based alternatives are also presented, using a combination of maximum likelihood and least squares methods. The thesis ends with a concluding discussion.

Acknowledgements

Primarily I would like to thank Dr Gavin Shaddick for his help, guidance and support over the last three years, without which this work would not have been possible. I would also like to thank him for his friendship during this time, which has made the transition to academic life easier. My thanks also go to Clare, whose love and support have helped me through the hard times at the expense of many other things. In addition I would like to thank my mum, dad and Rachel, who have always been there for me when I needed them. Finally I would like to thank all the friends I have made during the last three years for their help, especially James and those who play for the Venturers cricket team and the maths football team.

Declaration

The work presented in chapter five has been accepted for publication in Biometrics with the title *Time-varying coefficient models for the analysis of air pollution and health outcome data*, and is jointly authored with Gavin Shaddick. The same work has also been presented at the 17th annual meeting of the International Environmetrics Society (TIES) in Sweden in 2006, with the title *Estimating the effects of air pollution on public health using a time varying coefficient model.*

Contents

List of Figures

List of Tables

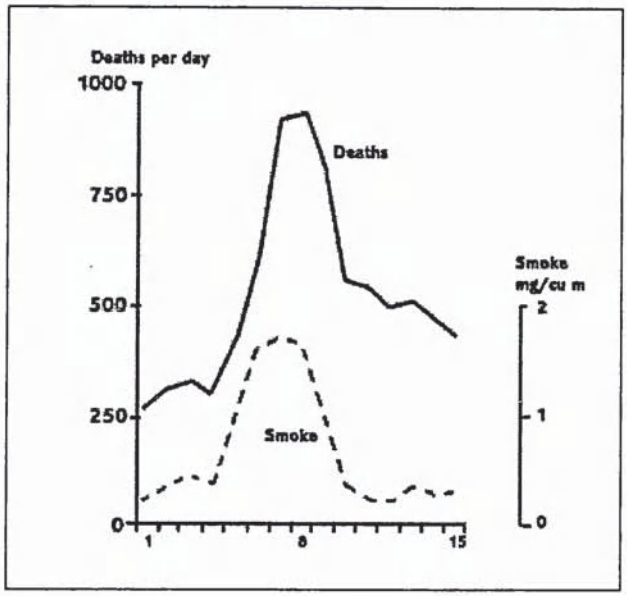
Chapter 1

Introduction

The association between air pollution exposure and mortality or morbidity has been a public health concern for over 700 years. As far back as 1272 King Edward I outlawed the burning of coal in London, causing a decrease in pollution levels across the city and an improvement in the health of the local population. However the subject has only become a global health issue in the last fifty years, primarily as a result of high air pollution episodes in the Meuse valley in 1930 (Firket (1936)), in Donora, Pennsylvania in 1948 (Ciocco and Thompson (1961)) and during the London smog of December 1952 (Ministry of Public Health (1954)). These episodes were caused by a combination of industrial pollution sources and adverse weather conditions, and resulted in large numbers of deaths among the surrounding populations. For example, as highlighted in Figure 1-1 the London smog was associated with more than 3000 excess deaths in December 1952 when compared with previous years. Although pollution levels in the last twenty years have been much lower than those witnessed in the above episodes, its relationship with both mortality and morbidity continues to be an active area of research. This research has led to a raft of legislation including the 'Clean Air Acts' (for example the Clean Air Act (1990) in the U.S. and the Clean Air Act (1993) in the UK), which have prohibited and regulated fixed and moving pollution sources. In addition, air quality standards such as the American National Ambient Air Quality Standards (1990) and the UK Air Quality Strategy (2000) have been implemented, which set legal levels for common pollutants and have brought the subject of air pollution into public focus.

The majority of air pollution studies have examined the effects of short-term (acute) exposure over a few days, rather than long-term (chronic) exposure over a number of years. The relationship between chronic exposure and adverse health is typically estimated using cohort studies, but their high costs mean such studies are relatively rare. Examples include the Six Cities Study (Dockery et al. (1993)) and the American Cancer Study (Pope et al. (1995) and Pope et al. (2002)) in the USA, and the Netherlands Cohort Study on Diet and Cancer (Hoek et al. (2002)) in Europe. In contrast numerous studies have

Figure 1-1: Mortality and pollution levels during the London Smog of December 1952 (taken from the Met Office website).



examined the relationship between acute exposure and mortality or morbidity, and can be classified into three broad types: case-crossover studies (see for example Neas et al. (1999)), panel studies (see for example Yu et al. (2000)), and time series studies (see for example Schwartz et al. (2001)). The first two of these use individual level data, allowing an exposure-response relationship to be estimated. However pollution related mortality or morbidity events are rare, and in order to produce conclusive results case-crossover and panel studies require data about a large number of individuals. Consequently these studies are costly to implement, and the majority of air pollution research is based on time series studies. These studies use aggregate level mortality or morbidity data, which describe the health of the population living within a geographical region rather than that of specific individuals. These data are also routinely available, meaning that time series studies are inexpensive and straightforward to implement. A second advantage of time series studies is that they are unlikely to be affected by individual level risk factors such as smoking habits, because the prevalence of such factors is likely to be constant over the period of study. However, the ecological nature of time series studies mean they estimate a group level association between air pollution exposure and health outcomes, which is much weaker than an individual exposure-response relationship (for further details see Wakefield and Salway (2001)). The remainder of this thesis focuses on time series studies, but for a more general review of air pollution research see Dominici et al. (2003) and Pope and Dockery (2006).

Time series studies are based on health, pollution and meteorological data from a geo-

graphical region such as a city. The health data comprise daily counts of mortality or morbidity drawn from the population living within this region, and a number of health classifications have been used in such studies. Examples include general categories such as total non-accidental mortality, and illness specific subclasses such as respiratory mortality and hospital admissions due to asthma. The air pollution data are obtained from a number of fixed site monitors, which are located throughout the region under study. These monitors measure ambient (background) pollution levels throughout the day, and a daily average is typically calculated at each site. Examples of meteorological covariates include temperature, dew-point temperature and humidity, which are also routinely measured by fixed site monitors in urban areas.

One of the first time series studies was carried out by Schwartz and Marcus (1990), who analysed data from Greater London using a normal linear model. However daily counts of mortality or morbidity may include small numbers, meaning that Poisson regression techniques such as linear (GLM, McCullagh and Nelder (1989)) and additive (GAM, Hastie and Tibshirani (1990)) models are appropriate. These models regress daily counts of mortality or morbidity against air pollution levels and a vector of covariates, the latter of which remove underlying trends, seasonal variation, over-dispersion and temporal correlation that routinely afflict health data of this type. These characteristics are typically induced by risk factors other than air pollution, such as meteorological conditions and influenza epidemics. Air pollution studies typically (see for example Schwartz (1991) and Spix et al. (1993)) analyse data from a small number of cities, and a number of statistical approaches have been adopted in the wider literature. This variation in statistical methods may be partly responsible for the considerable heterogeneity observed in the pollution-mortality associations estimated in the air pollution literature. A number of researchers have attempted to reduce this heterogeneity by implementing large multi-city studies, including Air Pollution and Health: A European Approach (APHEA, Schwartz et al. (1996) and Katsouyanni et al. (2001)) in Europe, and the National Morbidity, Mortality and Air Pollution Study (NMMAPS, Samet et al. (2000)) in the USA. These studies analyse data from each city using a standard protocol, which makes comparisons between multiple cities fairer.

In this thesis I extend the methods used to estimate the association between air pollution exposure and health, and compare their efficacy against those adopted by the majority of researchers. These developments are either less restrictive than the current methods, or provide the first evidence that the standard approaches to modelling these data are deficient. The work presented in this thesis is centered around three related themes, which incorporate both the air pollution and covariate components of the regression model. The first theme relates to the covariate component, focusing on how long-term trends, seasonal variation, over-dispersion and temporal correlation should be removed from the

health data. The majority of authors remove these factors using smooth functions of calendar time, and I compare this approach against a variety of discrete time processes. Such processes occur at discrete time intervals such as days, and the main example used in this thesis are autoregressive processes. The second theme focuses on the association between air pollution exposure and mortality, and in particular whether it evolves over time. Such temporal variation in the health risk associated with pollution exposure has only been investigated in a small number of studies, and the models presented here are more flexible than those previously used. The last theme focuses on air pollution exposure, and the impact that mis-estimating it has on the association with mortality. The majority of authors estimate pollution exposure with a simple average across all available monitors, and the accuracy of this approach is compared against less restrictive alternatives. These developments are investigated using Bayesian hierarchical models with analysis based on Markov chain monte carlo (MCMC) simulation. The first two themes are additionally implemented using likelihood based inference, and a comparison between these approaches is a further development in air pollution studies.

The remainder of this thesis is structured into six chapters, the first of which reviews the statistical methodology used in this thesis and the related literature. Chapter three critiques the air pollution and health literature, focusing on the standard approaches to analysis and their deficiencies. Chapter four analyses air pollution and health data using a dynamic generalised linear model, which allows the trend in daily mortality and the pollution-mortality relationship to be represented as autoregressive processes. Chapter five analyses these data with a time-varying coefficient model, which allows the pollution-mortality association to evolve over time. Chapter six compares the accuracy of using different estimates of pollution exposure, which include the simple average described above and a spatio-temporal pollution model. Finally chapter seven presents a concluding discussion, focusing on how this work impacts on future air pollution research. The Bayesian analyses presented in chapters four to six are based on Markov chain monte carlo algorithms written in the statistical language R (R 2.1.0, a Language and Environment (2005)), which were developed in conjunction with this thesis. They include a number of Metropolis-Hastings and Gibbs sampling steps, such as the conditional prior proposal method (Knorr-Held (1999)) and random walk updates. The remainder of this introduction describes the individual chapters in more detail.

Chapter two reviews the statistical methods used in this thesis, with particular emphasis on Bayesian analysis and regression methodology. Both Bayesian and likelihood approaches to analysis are outlined, including a review of estimation techniques such as maximum likelihood and Markov chain monte carlo simulation. A large proportion of this chapter outlines the regression models used in this thesis, including simple linear models as well as hierarchical alternatives. The latter include dynamic generalised linear models, varying

coefficient models, measurement error models and spatio-temporal models, which are the basis of the developments proposed in chapters four to six. The chapter ends with a review of model selection criteria. Chapter three critiques the air pollution and health literature, beginning by outlining the standard approaches to modelling these data. Particular emphasis is given to both air pollution levels and the covariates, the latter of which typically include trend components and meteorological conditions. The remainder of the chapter discusses the limitations of the standard approaches previously described.

Chapter four analyses air pollution and health data with a dynamic generalised linear model, which extends the standard model discussed in chapter three in two ways. Firstly underlying trends and temporal correlation present in the health data can be modelled by an autoregressive process, which contrasts with the standard approach of using smooth functions of calendar time. This is appropriate because the health data and the autoregressive processes sit in discrete time, whereas a smooth function has a continuous time support. Secondly the relationship between air pollution and mortality is allowed to evolve over time, which contrasts with the standard approaches of fixing it to be constant or modelling it as a dose-response curve. The approach adopted here allows any temporal variation in this relationship to be examined, which may be a long-term trend or exhibit periodic behaviour. Both these developments are investigated by applying a series of models to data from Greater London, with analysis implemented in both Bayesian and likelihood frameworks.

Chapter five extends the focus on time-varying pollution-mortality relationships examined in chapter four, by modelling them as a smooth function rather than an autoregressive process. A smooth function is adopted here because it forces any variation in the pollution-mortality relationship to evolve smoothly over time, a desirable characteristic that is not enforced by an autoregressive process. The temporal variation in this relationship is represented within a time-varying coefficient model, for which both Bayesian and likelihood implementations are presented. The efficacy of the model is assessed via a simulation study, where it is applied to data with different shaped time-varying relationships. The model is then used to analyse data from four U.S. cities; Cleveland, Detroit, Minneapolis and Pittsburgh.

Chapter six examines how mis-estimating pollution exposure affects the association with mortality, and begins by examining the accuracy of averaging the pollution measurements at each monitoring site. This is the approach adopted in the majority of air pollution studies, even though it ignores the possibility of measurement error in the observed data and spatial variation in the underlying pollution surface. As a result I propose two alternative estimates of pollution exposure, the first accounting for measurement error while the second additionally incorporates spatial variation. The last of these is based on a spatio-

temporal pollution model, which is the subject of much current research. The efficacy of each approach is assessed via simulation and case studies, where the estimates of pollution exposure and the association with mortality are of primary interest. In common with the rest of this thesis the models are implemented within a Bayesian framework, using Markov chain monte carlo simulation.

Chapter seven discusses the main results from this thesis and assesses its contribution to the wider literature. The limitations of this work are discussed, with possible extensions and future work outlined.

Chapter 2

Statistical Review

This chapter outlines the statistical theory used in this thesis, beginning with a comparison of likelihood and Bayesian approaches to analysis. The likelihood approach is the inferential framework used in the majority of air pollution studies (see for example Verhoeff et al. (1996), Hong et al. (1999) and Goldberg et al. (2001)), and a brief review of parameter estimation, confidence regions and hypothesis tests is given in section 2.1. However as data structures and models increase in complexity the Bayesian approach is becoming increasingly common, and is the statistical framework adopted throughout this thesis. It is introduced in section 2.2 where Bayes theorem, prior distributions and Markov chain monte carlo simulation are outlined. The remainder of this chapter discusses regression methodology, which is the central tool for estimating the association between air pollution exposure and mortality or morbidity. Sections 2.3 and 2.4 review generalised linear and additive models respectively, which are the standard approaches to modelling air pollution and mortality data in the related literature. Section 2.5 presents a general outline of hierarchical regression models, with the specific models used in chapters four to six reviewed in the next four sections. Section 2.6 presents a brief outline of dynamic generalised linear models (used in chapter four), while section 2.7 introduces the general class of varying coefficient models (used in chapter five). Sections 2.8 and 2.9 review measurement error and spatio-temporal models respectively, which are both applied in chapter six. Finally section 2.10 reviews model selection techniques in Bayesian and likelihood analyses, a collection of which are used in chapters four to six. For notational clarity vectors are denoted by bold type, matrices by upper case letters and scalar quantities by lower case letters.

2.1 Likelihood based methods

The likelihood or frequentist paradigm is based on a vector of observations $\mathbf{y} = (y_1, \ldots, y_n)_{n\times 1}$, which are assumed to come from a family of probability distributions f, indexed by unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)_{q\times 1}$. Each family of distributions is characterised by a probability density function or probability mass function $f(\mathbf{y}|\boldsymbol{\theta})$, which measures the

likelihood of observing different values of $\mathbf{y}$ for a given $\boldsymbol{\theta}$. However $\mathbf{y}$ is typically observed, and $f(\mathbf{y}|\boldsymbol{\theta})$ also measures how likely particular values of $\boldsymbol{\theta}$ are given this observed data. In this setting it is known as the likelihood function, and is denoted by $L(\boldsymbol{\theta}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\theta})$. If $\mathbf{y}$ is a vector of independent observations this simplifies to $L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$, the product of probability density/mass functions for each y_i. Many families of distributions are similar and have probability density/mass functions that can be expressed as

$$f(y_i|\theta) = \exp\left(\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i, \phi)\right),$$

for arbitrary functions (a, b, c), and univariate parameters (θ, ϕ). This is exponential family form and includes standard distributions such as Poisson, binomial, gamma, and Gaussian. The functions (a, b, c) determine the distributional family, while (θ, ϕ) specify the location and scale of the exact distribution within a given family. For an exponential family of distributions the mean, variance and Fisher information $(\upsilon(\theta))$ are given by

$$\mathbb{E}[y_i] = b'(\theta), \quad \text{Var}[y_i] = a(\phi)b''(\theta) \quad \text{and} \quad \upsilon(\theta) = \frac{b''(\theta)}{a(\phi)}.$$

The prime notation denotes differentiation, and for many families ϕ is fixed at one. In air pollution studies the health data are a vector of daily counts, meaning that each observation is typically modelled by a Poisson distribution. For this family of distributions θ represents the mean, ϕ is fixed at one, and the mean and variance are equal. However the last of these assumptions is restrictive, and many sets of health data are overdispersed, exhibiting more variation than is assumed by the Poisson distribution. In this situation the Poisson likelihood can be replaced by a quasi-likelihood, which is an approximation that incorporates additional variation by allowing ϕ to be greater than one. A brief outline of the quasi-likelihood approach is given in chapter three, along with a review of the data used in air pollution studies.

2.1.1 Inference

After a family of probability distributions has been specified for $\mathbf{y}$, inference takes the form of point estimates, confidence regions and hypothesis tests, each of which are outlined below.

Point estimates

A point estimate is the value of $\boldsymbol{\theta}$ that is most supported by the observed data, and can be calculated using a number of generic methods. Two of the most common are maximum likelihood and least squares, both of which are outlined below. An alternative technique is the method of moments (Hinkley et al. (1991)), and comprehensive reviews of point estimation are given by Silvey (1975) and Bates and Watts (1988). The maximum

likelihood estimator (MLE) is the value of θ that maximises the likelihood or log likelihood functions, the latter of which is denoted by $\mathcal{L}(\theta|\mathbf{y}) = \log[L(\theta|\mathbf{y})]$. It can be calculated by solving the vector equation

$$\mathcal{L}'(\theta|\mathbf{y}) = \frac{d\mathcal{L}(\theta|\mathbf{y})}{d\theta} = \mathbf{0}, \tag{2.1}$$

providing the solution is a local maximum. The $q \times 1$ vector of first derivatives $\mathcal{L}'(\theta|\mathbf{y})$ is called the score function, while $\mathcal{L}'(\theta|\mathbf{y}) = \mathbf{0}$ is known as the score equation. If (2.1) is sufficiently complex it cannot be solved analytically, and a numerical method is required. One example is Newton's method (for further details see Lange (1998)), which approximates the score function by a first order Taylor series leading to the iterative solution

$$\theta^{(j+1)} = \theta^{(j)} - \mathcal{L}''(\theta^{(j)}|\mathbf{y})^{-1}\mathcal{L}'(\theta^{(j)}|\mathbf{y}). \tag{2.2}$$

A slight modification is the method of scoring, which is used for generalised linear models and replaces the Hessian matrix $\mathcal{L}''(\theta^{(j)}|\mathbf{y})$ with the Fisher information matrix $-\mathbb{E}\left[\mathcal{L}''(\theta^{(j)}|\mathbf{y})\right]$. An alternative to maximum likelihood is generalised least squares (GLS), which only uses the mean and variance of $\mathbf{y}$ rather than its full distribution. Denoting the mean and variance of $\mathbf{y}$ by $\mu(\theta)_{n\times 1}$ and $\Sigma(\theta)_{n\times n}$ respectively, the generalised least squares estimator is the value of θ that minimises

$$S(\theta|\mathbf{y}) = (\mathbf{y} - \mu(\theta))^{\mathrm{T}}\Sigma(\theta)^{-1}(\mathbf{y} - \mu(\theta)),$$

the generalised sum of squares. In the special case that $\mu(\theta) = X\theta$ and $\Sigma(\theta) = \Sigma$, the GLS estimator has the closed form

$$\tilde{\theta} = (X^{\mathrm{T}}\Sigma^{-1}X)^{-1}X^{\mathrm{T}}\Sigma^{-1}\mathbf{y}. \tag{2.3}$$

However in general $\tilde{\theta}$ has no closed form expression, and must be estimated using a numerical method such as those described above. If $\Sigma(\theta)$ is diagonal with unequal variances GLS is called weighted least squares, while the further simplification $\Sigma = \sigma^2 I$ is known as ordinary least squares.

Confidence regions and hypothesis tests

After θ has been estimated inferential tools such as confidence regions and hypothesis tests can be calculated, to give a better understanding of the underlying process. A confidence region for $\theta_{q\times 1}$ is a subspace of $\mathbb{R}^q$ in which its true value is expected to lie. The likelihood framework is based on the idea of repeated sampling, in which an infinite number of hypothetical data sets $\mathbf{y}^{(i)}$ can be generated. Each of these can be used to construct a confidence region, and a 95% confidence region is calculated so that 95% of these hypothetical regions contain the true value of θ. For a univariate parameter the region simplifies to an interval

of the form $(\hat{\theta} \pm D)$, in which $\hat{\theta}$ is a point estimate and D depends on the distribution and standard error of $\hat{\theta}$. In contrast a hypothesis test makes a particular statement about θ (called H_0), and uses the data to assess the level of support for that statement. A test statistic T is calculated from the data, and the probability of obtaining a result as or more extreme than T assuming H_0 is true is calculated. This probability is called the p-value, and if it is too small, typically less than 5%, the original statement is deemed to be untrue.

Confidence regions and hypothesis tests are commonly used in air pollution studies, both in the model building and inferential stages. Model building is typically based on hypothesis tests, which are used to determine the set of covariates to include in the final model. Inference focuses on the estimated relationship between air pollution and health, which is typically presented as a parameter estimate and confidence interval. These are typically calculated on the relative risk scale (see chapter 3.2), and if the interval does not contain one there is evidence that air pollution is harmful to human health.

2.2 Bayesian methods

In common with the likelihood approach the building blocks of a Bayesian analysis are the data $\mathbf{y}$ and the parameter vector θ, and uncertainty in the former is described by $f(\mathbf{y}|\theta)$. However in a Bayesian analysis the uncertainty about θ before $\mathbf{y}$ has been observed is also quantified. This uncertainty is expressed as a probability distribution $f(\theta)$, which is called a prior and is discussed in more detail in section 2.2.1. The aim of a Bayesian analysis is to learn about θ, which is achieved by determining its posterior distribution conditional on the observed data. This distribution is given by

$$f(\theta|\mathbf{y}) = \frac{f(\theta, \mathbf{y})}{f(\mathbf{y})} = \frac{f(\theta)f(\mathbf{y}|\theta)}{f(\mathbf{y})},$$

which is obtained from an application of Bayes theorem. The denominator $f(\mathbf{y})$ is the marginal distribution of the data, and is calculated as $f(\mathbf{y}) = \sum_{\theta} f(\theta)f(\mathbf{y}|\theta)$ if θ is discrete, and $f(\mathbf{y}) = \int_{\theta} f(\theta)f(\mathbf{y}|\theta)d\theta$ if θ is continuous. If θ is multivariate $f(\mathbf{y})$ is based on multidimensional integrals, which can be analytically intractable and computationally expensive to estimate. As a result Bayesian analysis is typically based on the unnormalised posterior distribution

$$f(\theta|\mathbf{y}) \propto f(\theta)f(\mathbf{y}|\theta),$$

which is the product of the likelihood function and the prior distribution. This unnormalised posterior can be used instead of $f(\theta|\mathbf{y})$ because $f(\mathbf{y})$ has no dependence on θ. The ability to summarise θ via its posterior distribution is a major advantage of the Bayesian framework, because $f(\theta|\mathbf{y})$ contains more information about θ than is typically obtained

from a likelihood analysis. In a likelihood setting inference is based on point estimates and confidence regions, which have parallels in the Bayesian paradigm. The posterior mean ($\mathbb{E}(\boldsymbol{\theta}|\mathbf{y}) = \int_{\boldsymbol{\theta}} \boldsymbol{\theta} f(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$) and median are typically quoted as point estimates, while a Bayesian 95% confidence region $\mathcal{R}$, satisfies $\mathbb{P}(\boldsymbol{\theta} \in \mathcal{R}|\mathbf{y}) = 95\%$. This is called a credible region, and unlike its likelihood counterpart the probability of $\boldsymbol{\theta}$ lying in $\mathcal{R}$ is 95%. The remainder of this section discusses prior and posterior distributions, while a more general review of Bayesian methods is given by Gelman et al. (2003).

2.2.1 Prior distributions

A prior distribution represents the information about $\boldsymbol{\theta}$ before any data are observed, which may be prior ignorance or come from previous studies of similar data. It is typically represented by a standard probability distribution, which depends on a vector of hyper-parameters that may or may not be known. The form of $f(\boldsymbol{\theta})$ depends on the problem, and may comprise a single multivariate distribution, a product of independent univariate distributions, or a combination of marginal and conditional distributions. Two classes of prior distribution used in this thesis are conjugate and non-informative, both of which are briefly outlined. A conjugate prior is one that has the same functional form as the likelihood, meaning that the full conditional posterior distribution will come from the same family of distributions as the prior. For example consider a single Gaussian observation $y \sim \mathrm{N}(\mu, \sigma^2)$, which has unknown parameters μ and σ^2. The conjugate priors are $\mathrm{N}(\mu_0, \sigma_0)$ and inverse-gamma(e, f) distributions respectively, which result in the following full conditional posterior distributions:

$$\begin{aligned} f(\mu|\sigma^2, y) &\sim \mathrm{N}\left(\left[\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}\right]^{-1}\left[\frac{y}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right], \left[\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}\right]^{-1}\right), \\ f(\sigma^2|\mu, y) &\sim \text{Inverse-Gamma}\left(e + \frac{1}{2}, \frac{1}{2}(y-\mu)^2\right). \end{aligned}$$

Conjugate priors are attractive because the posterior is from a standard family of distributions, making its computation relatively straightforward (see section 2.2.2). However such priors cannot always be used, because they might not be available or may not be a suitable specification of prior knowledge. Alternatively prior ignorance can be expressed using a non-informative (flat) distribution, which gives fairly even support to a wide range of values. Non-informative priors can be a standard distribution with a large variance or be improper, the latter of which does not have a finite density, for example a uniform$[-\infty, \infty]$ distribution. Whilst initially appealing improper priors should be used with caution, as the resulting posterior distribution may also be improper making analysis impossible. As a result, the majority of researchers specify prior ignorance using a proper prior with large variance, with two common choices being Gaussian (if the parameter space is equal to the

whole real line), or inverse-gamma (for parameters that are strictly positive) distributions. The latter of these is a common choice for variance parameters, and is used extensively in this thesis.

2.2.2 Inference

The posterior distribution can be calculated using a variety of techniques, the choice of which depends on the complexity of the combination of the likelihood and the prior. In very simple problems with conjugate priors the posterior distribution can be obtained analytically, because it comes from a standard family of distributions. However in most situations this is not possible, and the posterior distribution must be estimated by simulation. In this situation a large number of samples are drawn from $f(\boldsymbol{\theta}|\mathbf{y})$, which can be used to estimate quantities of interest such as the posterior mean. These samples can be generated using a number of techniques, the simplest of which are direct methods such as inversion or rejection sampling (for a review see Ripley (1987)). These methods are preferable over more complex alternatives, because they generate independent samples and require little computational effort. However the majority of problems are too complex for direct methods, and approximate iterative techniques are required. The most popular of these is Markov chain monte carlo (MCMC) simulation, which is used extensively for the Bayesian analyses presented in this thesis. A brief review of MCMC is given below, while more comprehensive treatments are given by Gilks et al. (1996), Gelman et al. (2003) and Gamerman and Lopes (2006).

Markov chain monte carlo simulation

Markov chain monte carlo simulation is based on a Markov chain, $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(3)} \ldots\}$, whose target distribution is the joint posterior $f(\boldsymbol{\theta}|\mathbf{y})$. The chain is initialised by a starting value $\boldsymbol{\theta}^{(0)}$, and is run until it has converged to its target distribution. Convergence can be assessed using the objective criteria proposed by Gelman and Rubin (1992), and the initial period of non-convergence is known as 'burn-in'. After convergence has been reached the Markov chain is sampling from the posterior distribution, and can generate as many samples as are required. Markov chain simulation is one of the few approaches that can be used when the statistical model is sufficiently complex, but is inferior to direct methods in a number of ways. For example unlike direct methods successive simulations from a Markov chain are not independent, meaning that a sample provides less information about the posterior distribution than an independent sample of the same size. In addition it is not possible to determine whether the Markov chain has covered the entire posterior distribution, leaving the possibility that it may have missed an area of non-negligible posterior probability. However, these problems can be partially alleviated by using multiple Markov chains initialised at dispersed points in the sample space, which reduces the likelihood that areas of non-negligible posterior probability have been missed. In addition the

samples could be thinned, which is the process of only keeping every kth sample. This reduces the correlation in successive samples but increases the computational burden, because a large number of samples are required as the majority are discarded.

Markov chain simulation can be implemented using the Metropolis-Hastings algorithm (Metropolis et al. (1953) and Hastings (1970)) and the Gibbs sampler (Smith and Roberts (1993)), both of which have a target distribution equal to the posterior distribution of interest. These algorithms partition the parameter vector into $\theta = (\theta_1, \ldots, \theta_d)$, and a single iteration sequentially updates each of the d blocks in turn. The most general of these algorithms is Metropolis-Hastings which is described below.

Algorithm 1. *Metropolis-Hastings*

1. *Draw a starting point for the Markov chain ensuring that its posterior probability is positive, that is* $f(\theta^{(0)}|\mathbf{y}) > 0$.

2. *For each iteration* $j = 1, 2, \ldots$ *carry out steps (a) and (b) for each of the* d *sub-vectors* $\theta_1, \ldots, \theta_d$. *For sub-vector* θ_k:

   (a) *Generate a possible sample* θ_k^* *from a proposal distribution* $q(\theta_k^{(j)}, \theta_k^*)$, *that is based on the current value of the chain.*

   (b) *Accept* θ_k^* *as the next iteration, that is set* $\theta_k^{(j+1)} = \theta_k^*$, *with probability*

$$r = min\left\{1\ ,\ \frac{f(\theta^*|\mathbf{y})q(\theta^*, \theta^{(j)})}{f(\theta^{(j)}|\mathbf{y})q(\theta^{(j)}, \theta^*)}\right\},$$

   and reject it (that is set $\theta_k^{(j+1)} = \theta_k^{(j)}$*) with probability* $1 - r$. *In calculating the acceptance probability* θ^* *and* $\theta^{(j)}$ *are identical except at sub-vector* k.

The Gibbs sampler is a special case of this algorithm, in which the proposal distribution is given by $q(\theta_k^{(j)}, \theta_k^*) = f(\theta_k|\theta_{-k}^{(j)}, \mathbf{y})$, the full conditional distribution of θ_k (here $\theta_{-k} = (\theta_1, \ldots, \theta_{k-1}, \theta_{k+1}, \ldots, \theta_d)$). For this algorithm the acceptance probability simplifies to one, which effectively removes the accept or reject stage. Markov chain simulation is complex to implement, and the results can be affected by the choice of starting distribution, partition of the parameter vector, proposal distribution, and the desired acceptance rates, all of which are briefly discussed.

Starting distribution

The starting distribution must have the same sample space as the posterior (for example variance parameters should not be drawn from distributions which allow negative values), and is typically chosen to be an overdispersed version of the prior. This strategy allows the starting locations of multiple chains to be highly dispersed, which increases the likelihood

of covering all areas of non-negligible posterior probability.

Partition of θ

The parameter vector can be sampled 'all at once', in blocks or singularly, and while the former produces better mixing the latter results in higher acceptance rates. An appropriate partition of θ depends on the problem, and parameters with similar characteristics are typically sampled in a single block. Each block can be updated using a separate proposal distribution, meaning that an algorithm can contain a mixture of Metropolis-Hastings and Gibbs sampling steps.

Proposal distribution

An adequate proposal distribution depends on the problem, and if the full conditional of θ_k is from a standard family Gibbs sampling is typically preferred. If not a Metropolis-Hastings algorithm can be implemented, although the choice of proposal distribution has a large impact on the convergence and acceptance rate of the algorithm. Proposal distributions can often be problem specific, although a Gaussian random walk is widely used in a number of situations. This proposal distribution generates θ_k^* from $\mathrm{N}(\theta_k^{(j)}, \sigma^2 I)$, which is equal to the current value $\theta_k^{(j)}$ plus random error. The variance of this distribution controls the acceptance rate of the algorithm, which increases as $\sigma^2 \to 0$. An additional advantage of this proposal is that it is symmetric in $(\theta_k^{(j)}, \theta_k^*)$, meaning that its acceptance probability simplifies to $f(\theta_k^*|\mathbf{y})/f(\theta_k^{(j)}|\mathbf{y})$.

Acceptance rate

The acceptance rate is determined by the variance of the proposal distribution, with large variances allowing bigger moves around the sample space at the expense of more rejections. Some proposal distributions include one or more tuning parameters, that can be altered to achieve the desired acceptance rate. This desired rate depends on a number of factors, including the stage of the simulation and the choice of proposal distribution. For example the desired acceptance rate for a Gaussian random walk proposal is known to be 23% (for details see Fahrmeir and Tutz (2001)).

2.3 Generalised linear models

Regression models estimate the relationship between a set of response data $\mathbf{y} = (y_1, \ldots, y_n)_{n\times 1}$, and a matrix of q covariates $X = (\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_n^{\mathrm{T}})_{n\times q}$, where $\mathbf{x}_i^{\mathrm{T}} = (x_{i1}, \ldots, x_{iq})_{1\times q}$ denotes the realisations relating to observation i. In air pollution studies analysis is typically based on Poisson linear or additive models, the first of which is reviewed here while the latter is outlined in the next section. In a generalised linear model (GLM, Nelder and Wedderburn (1972)) each y_i is assumed to be an independent observation from an exponential family distribution f. The model is given by

$$\begin{aligned} y_i &\sim f(y_i|\mu_i) \quad \text{for} \quad i = 1, \ldots, n, \\ g(\mu_i) &= \eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \end{aligned} \tag{2.4}$$

where μ_i denotes the expected value of y_i. The covariates are multiplied by a vector of unknown regression parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)_{q\times 1}$, which represent the relationship between each covariate and the response. The linear combination of all covariates is called the linear predictor (η_i), and is related to the expected value by a known invertible link function g. The only unknown quantity in this model is $\boldsymbol{\beta}$, and an outline of likelihood and Bayesian estimation is given below. For comprehensive reviews of generalised linear models see McCullagh and Nelder (1989) and Dobson (1990).

2.3.1 Likelihood based estimation

The maximum likelihood estimate of $\boldsymbol{\beta}$ is obtained by solving the vector valued score equation (2.1), which for this model is given by

$$\mathcal{L}'(\boldsymbol{\beta}|\mathbf{y}) = X^{\mathrm{T}}W(\boldsymbol{\beta})M(\boldsymbol{\beta})(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta})) = \mathbf{0}.$$

In the equation above $\boldsymbol{\mu}(\boldsymbol{\beta}) = (\mu_1, \ldots, \mu_n)^{\mathrm{T}}_{n\times 1}$ is the vector of expected values, while $M(\boldsymbol{\beta})$ and $W(\boldsymbol{\beta})$ are diagonal $n \times n$ matrices with elements

$$\begin{aligned} m_{ii} &= \frac{\partial \eta_i}{\partial \mu_i} \text{ and} \\ w_{ii} &= \frac{1}{\mathrm{Var}[y_i]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2, \end{aligned} \tag{2.5}$$

respectively. The score equation is non-linear in $\boldsymbol{\beta}$ and can be solved using Newton's method (equation 2.2)) with a Fisher scoring step. The jth estimate of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}^{(j+1)} = \boldsymbol{\beta}^{(j)} + \left[X^{\mathrm{T}}W(\boldsymbol{\beta})X\right]^{-1} X^{\mathrm{T}}W(\boldsymbol{\beta})M(\boldsymbol{\beta})[\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta})], \tag{2.6}$$

which is iterated until convergence of successive estimates. The algorithm is initialised by a starting value $\boldsymbol{\beta}^{(0)}$, which is typically a vector of zeros or based on estimates from a simpler model. The Fisher information matrix is given by $X^{\mathrm{T}}W(\boldsymbol{\beta})X$, and (2.6) can be re-written as

$$\boldsymbol{\beta}^{(j+1)} = \left[X^{\mathrm{T}}W(\boldsymbol{\beta}^{(j)})X\right]^{-1} X^{\mathrm{T}}W(\boldsymbol{\beta}^{(j)})\tilde{\mathbf{y}}(\boldsymbol{\beta}^{(j)}),$$

a generalised least squares step. In the above equation $\tilde{\mathbf{y}}(\boldsymbol{\beta}^{(j)}) = \boldsymbol{\eta}(\boldsymbol{\beta}^{(j)}) + M(\boldsymbol{\beta}^{(j)})[\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta}^{(j)})]$ is a vector of 'working observations' and $W(\boldsymbol{\beta}^{(j)})$ is the 'working covariance ma-

trix', both of which occur frequently in the next few sections. The maximum likelihood estimator is asymptotically unbiased, and has an approximate Gaussian distribution given by $\hat{\beta} \overset{\text{approx}}{\sim} \mathrm{N}\left(\beta, \left[X^{\mathrm{T}} W(\hat{\beta}) X\right]^{-1}\right)$, which forms the basis of confidence intervals and hypothesis tests.

2.3.2 Bayesian estimation

The Bayesian generalised linear model consists of equation (2.4) and a prior $f(\beta)$, the latter of which is typically a multivariate Gaussian distribution. A variety of MCMC schemes have been proposed to estimate β, which include Gibbs sampling and Metropolis-Hastings algorithms. Dellaportas and Smith (1993) propose a univariate Gibbs sampling algorithm, which updates each regression parameter in turn using adaptive rejection sampling (Gilks and Wild (1992)). However this becomes computationally expensive as the number of covariates increase, and a Metropolis-Hastings step that updates β in blocks or 'all at once' is preferable. Fahrmeir and Tutz (2001) suggest using random walk or Fisher scoring proposals, the former being preferable because of the speed of variable generation and the existence of a tuning parameter. This is the approach adopted in chapters four to six when sampling regression parameters of this type. The parameters are updated in blocks $\beta_{r,s} = (\beta_r, \ldots . \beta_s)$, which is a sensible intermediate strategy between the high acceptance rates that result from univariate sampling and the improved mixing that comes from sampling β in one step.

2.3.3 Regression splines

Generalised linear models force a linear relationship between each covariate and $g(\mu_i)$, the size of which is represented by β. However in practice these relationships may be highly non-linear, and a less restrictive approach is to replace $x_{ij}\beta_j$ with a smooth function $S_j(x_{ij}|\lambda_j, \theta)$ whose shape is determined from the data. Such functions are traditionally estimated within a generalised additive model (see section 2.4) using non-parametric methods, but can also be implemented using parametric regression splines within a generalised linear model. Regression splines are less flexible than non-parametric alternatives, but their parametric nature makes them straightforward to implement within a Bayesian analysis. A regression spline is a piecewise polynomial that is subject to continuity and derivative constraints at a number of fixed points. The range of the covariate is partitioned into $k+1$ intervals by a set of k knots $\mathbf{r} = (r_1, \ldots, r_k)_{k\times 1}$, which are typically equally spaced throughout the data. A spline consists of polynomials of degree d between each pair of knots, which are constrained to be continuous and have $d-1$ continuous derivatives at the knots. Although the choice of d depends on the specific problem, natural cubic splines are commonly used because they are visually smooth and linear beyond the end knots which precludes erratic tail behaviour. A regression spline is represented by a linear combination of k basis functions, meaning that its value at observation i is given by

$$S_j(x_{ij}|\lambda, \theta = (\theta_1, \ldots, \theta_k)) = \sum_{l=1}^{k} B_l(x_{ij})\theta_l.$$

Regression splines fit into the generalised linear model structure because $\sum_{l=1}^{k} B_l(x_{ij})\theta_l$ is a linear combination of known basis functions just like $\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$. The complete matrix of basis functions is given by $B = (\mathbf{B}_1^{\mathrm{T}}, \ldots, \mathbf{B}_n^{\mathrm{T}})$, where $\mathbf{B}_i^{\mathrm{T}} = (B_1(x_{ij}), \ldots, B_k(x_{ij}))_{1\times k}$ relate to observation i and $B\boldsymbol{\theta}$ represents the complete non-linear relationship. The basis matrix can be represented by a variety of functions, two of which are truncated polynomials and B splines. The truncated power basis of degree g is given by $\mathbf{B}_i^{\mathrm{T}} = (1, x_{ij}, x_{ij}^2, \ldots, x_{ij}^g, (x_{ij} - r_1)_+^g, \ldots, (x_{ij} - r_k)_+^g)$, where

$$(x_{ij} - r_l)_+^g = \begin{cases} (x_{ij} - r_l)^g & \text{if } (x_{ij} - r_l) \text{ is positive} \\ 0 & \text{otherwise} \end{cases}.$$

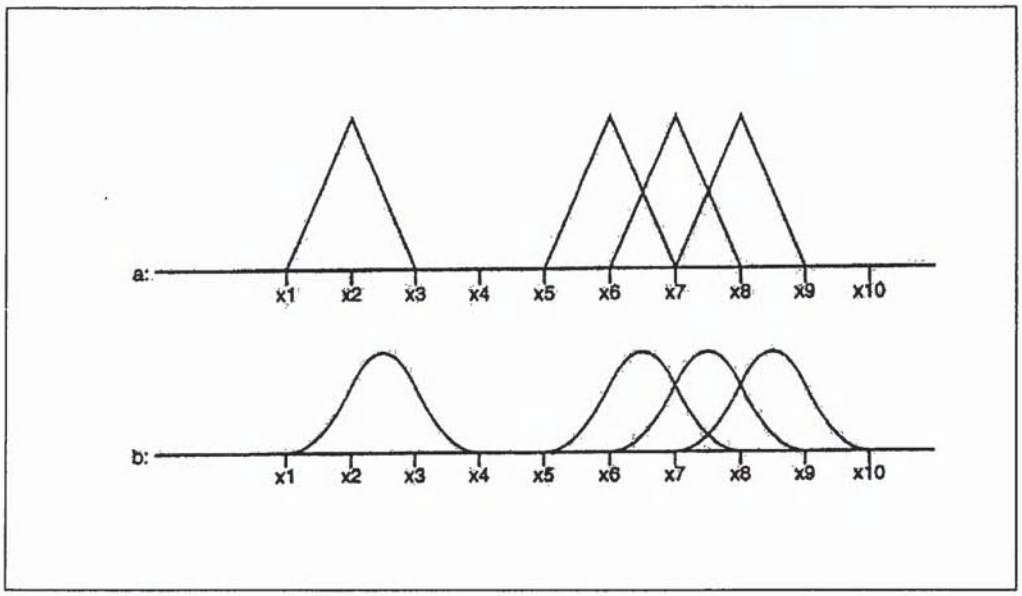
In contrast the B-spline basis consists of local polynomials (see Figure 2-1), which are zero over all but a few knots. As a result it is more numerically stable than the truncated power basis, and is typically preferred. The shape and smoothness of the spline depend on $\boldsymbol{\theta}$ and the smoothness parameter λ, the latter of which can be represented in numerous ways. The most common of these is to space the knots equally throughout the data, and specify the smoothness by k the number of knots (which is equal to λ in this situation). Alternatively a penalised approach can be adopted, which uses an overly large number of basis functions and penalises excess curvature in the estimate using a penalty term. In either case λ can be estimated using a data driven criterion, such as generalised cross validation (GCV) or Akaikes information criteria (AIC). Further details of regression splines can be found in Hastie and Tibshirani (1990) and Ruppert et al. (2003).

2.4 Generalised additive models

Generalised additive models (Hastie and Tibshirani (1986)) extend generalised linear models by replacing the linear relationship $x_{ij}\beta_j$ with a smooth function $S(x_{ij}|\lambda)$, whose shape is estimated from the data. Although this can be achieved using regression splines as described above, generalised additive models estimate these functions non-parametrically, meaning that they are more flexible and have the ability to exhibit greater non-linearity. In common with the linear model each y_i is assumed to be an independent observation from an exponential family distribution f. The model is given by

$$\begin{aligned} y_i &\sim f(y_i|\mu_i) \quad \text{for} \quad i = 1, \ldots, n, \\ g(\mu_i) &= \eta_i = \beta_0 + \sum_{j=1}^{q} S_j(x_{ij}|\lambda_j), \end{aligned} \tag{2.7}$$

Figure 2-1: Examples of B-spline basis functions of degrees one (a) and two (b) (taken from Eilers and Marx (1996)).



where the relationship between $g(\mu_i)$ and each covariate is represented by a smooth function. In common with regression splines the smoothness of each function is controlled by a single parameter λ_j, which can be estimated by optimising a data driven criterion. An intercept term is included so that each S_j can have mean zero, which removes non-identifiability problems that would result from estimating separate intercept parameters for each function. The remainder of this section outlines the methods used to represent non-parametric smooth functions and briefly reviews likelihood and Bayesian estimation of $(\beta_0, S_1, \ldots, S_q)$. For a comprehensive review of smooth functions and generalised additive models see Hastie and Tibshirani (1990).

2.4.1 Non-parametric smooth functions

Smooth functions can be estimated using a number of non-parametric methods, two of which are described below. For a general set of data $(y_1, \ldots, y_n)$ and $(x_1, \ldots, x_n)$, a kernel smoother estimates the underlying trend $S(x_i|\lambda)$ by

$$S(x_i|\lambda) = \frac{\sum_{l=1}^{n} \xi\left(\frac{|x_i - x_l|}{\lambda}\right) y_l}{\sum_{l=1}^{n} \xi\left(\frac{|x_i - x_l|}{\lambda}\right)},$$

a weighted average of $(y_1, \ldots, y_n)$. The kernel function ξ is decreasing in $|i|$, and common examples include the Gaussian and Epanechnikov kernels (for a review see Hastie and Tibshirani (1990)). The simplest example is a five point moving average, which estimates

$S(x_i|\lambda)$ by the average of the five y_l that have x_l closest to x_i. A further modification was proposed by Cleveland (1979), who estimates $S(x_i|\lambda)$ by locally weighted least squares regression. An alternative to kernel smoothers are smoothing splines, which are the solution to an optimisation problem rather than being constructed explicitly. A smoothing spline minimises the penalised sum of squares

$$\sum_{i=1}^{n}(y_i - S(x_i|\lambda))^2 + \lambda \int_a^b (S''(i|\lambda))^2 di, \tag{2.8}$$

subject to S having two continuous derivatives. The first term measures the closeness to the data while the second penalises curvature in S, where the limits on the penalty term are chosen to cover the range of the data. The smoothing parameter λ controls the size of the penalty term, and (2.8) is minimised by a natural cubic spline with knots at $(x_1, \ldots, x_n)$.

2.4.2 Likelihood based estimation

Likelihood based estimation is implemented using an iteratively re-weighted backfitting algorithm, similar to the iteratively re-weighted least squares procedure described in the previous section. The unknown quantities are an intercept term β_0 and q smooth functions $(S_1, \ldots, S_q)$, which are assigned reference starting values such as $\beta_0^{(0)} = g(\sum_{i=1}^n y_i/n)$ and $S_1^0 = \ldots = S_q^0 = 0$. For each iteration working observations and weights $\{\tilde{y}(\beta^{(j)}), w_{ii}(\beta^{(j)})\}_{i=1}^n$ are calculated as described in the previous section, and the unknown quantities are updated by applying a weighted backfitting algorithm to these data. The backfitting algorithm sequentially updates each S_i in turn, and replaces the generalised least squares step from the linear model. The smoothing parameters can be estimated by optimising a data driven criterion, and further details are given by Hastie and Tibshirani (1990).

2.4.3 Bayesian estimation

In the Bayesian paradigm generalised additive models are relatively new, and $(\beta_0, S_1, \ldots, S_q)$ can be estimated by MCMC simulation. Hastie and Tibshirani (2000) propose a Bayesian backfitting algorithm for this purpose, in which the optimisation of each S_j is replaced by posterior simulation. They use a Metropolis-Hastings algorithm where the proposal distribution for S_j is its full conditional distribution given the working observations and weights. The smoothing parameters can be estimated within the MCMC algorithm or using an empirical Bayes approach, the latter of which uses maximum likelihood methods. An alternative simulation algorithm has been suggested by Fahrmeir and Lang (2001), who use Markov random field priors within a generalised additive mixed model (GLMM) framework.

2.5 Hierarchical regression models

Hierarchical or multilevel models are a generic class of regression models that incorporate variation at multiple levels. For temporal data such as $\mathbf{y}$ this additional variation may be in the regression parameters (for example in a dynamic generalised linear model), covariates (for example in a measurement error model), or come from stochastic elements such as a set of random effects or an autoregressive process. The analyses presented in chapters four to six are based on hierarchical models of this type, using both Bayesian and likelihood approaches to estimation. The Bayesian approach provides the natural framework in which to analyse hierarchical data and models, because it incorporates multiple levels of variation in a straightforward manner. Hierarchical structures typically contain a number of conditional independence assumptions, meaning that the full conditional distribution for a block of parameters is typically much simpler than the full posterior distribution. The next four sections describes the hierarchical models used in chapters four to six, including dynamic generalised linear models, varying coefficient models, measurement error models, and spatio-temporal models.

2.6 Dynamic generalised linear models

Dynamic generalised linear models (DGLM, West et al. (1985)) are a generalisation of Gaussian state space models that were first introduced by Kalman (1960). They extend generalised linear models by allowing the regression parameters to evolve over time via an autoregressive process, which enables any variation in the covariate-response relationships to be captured. The response data come from an exponential family distribution f, whose mean is related to a set of covariates by a known and invertible link function g. The model is given by

$$\begin{aligned} y_i &\sim f(y_i|\mu_i) \quad \text{for} \quad i = 1, \ldots, n, \\ g(\mu_i) &= \eta_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_i, \\ \boldsymbol{\beta}_i &\sim \mathrm{N}(\Phi \boldsymbol{\beta}_{i-1}, \Sigma_\beta), \\ \boldsymbol{\beta}_0 &\sim \mathrm{N}(\boldsymbol{\mu}_0, \Sigma_0), \end{aligned} \tag{2.9}$$

where the regression parameters are collectively denoted by $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_n)_{(n+1)q \times 1}$, while $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{iq})_{q \times 1}$ represent those relating to observation i. The evolution of the autoregressive parameter process is determined by Φ and Σ_β, which are the $q \times q$ state and variance matrices respectively. The process is initialised by $\boldsymbol{\beta}_0$, which is assigned a multivariate Gaussian distribution with hyperparameters $(\boldsymbol{\mu}_0, \Sigma_0)$. A generalised linear model can be recovered by setting Φ equal to the identity matrix and Σ_β equal to a matrix of zeros. Comprehensive reviews of dynamic generalised linear models are given

by West and Harrison (1999) and Fahrmeir and Tutz (2001), while outlines of estimation and autoregressive processes are given below.

2.6.1 Estimation

There are numerous methods for estimating $(\boldsymbol{\beta}, \Phi, \Sigma_\beta, \mu_0, \Sigma_0)$, and only a brief review is given here. The three most common are the fully Bayesian approach (adopted in chapter four), the semi-Bayesian implementation of West et al. (1985) and the likelihood based methods of Fahrmeir and co-workers (also used in chapter four). The Bayesian approach is discussed in detail in chapter four, where a complete MCMC algorithm for model (2.9) is presented. The semi-Bayesian approach of West et al. (1985) is an approximate analysis, that relaxes the normality of the parameter process and assumes conjugacy between the data model and the autoregressive parameter model. They use linear Bayes methods to estimate the conditional moments of $\boldsymbol{\beta}_i$, and circumvent estimation of Σ_β using the discount method (Ameen and Harrison (1985)). In contrast the likelihood based approach estimates $\boldsymbol{\beta}$ by its posterior mode, and has been implemented using a number of algorithms proposed by Fahrmeir and co-workers (see Fahrmeir and Kaufmann (1991), Fahrmeir (1992) and Fahrmeir and Wagenpfeil (1997)). The latter of these is outlined below, because it was used in the only air pollution study to adopt a dynamic model. This study was presented by Chiogna and Gaetan (2002), and the Bayesian models in chapter four are compared against this likelihood based alternative. Fahrmeir and Wagenpfeil (1997) propose an iteratively re-weighted version of the Kalman filter and smoother, which is similar to the estimation algorithm used for generalised linear models. The parameters are assigned starting values $\boldsymbol{\beta}^{(0)}$, which are typically generated with the generalised extended kalman filter and smoother (Fahrmeir (1992)) or a generalised linear model, the latter of which forces each covariate-response relationship to be constant. At each iteration working observations and weights are calculated as described in section 2.3, and $\boldsymbol{\beta}$ is updated by applying the Kalman filter and smoother (Algorithm 2 below) to these pseudo data. The algorithm is iterated until convergence of successive estimates, while the hyperparameters $(\Phi, \Sigma_\beta, \mu_0, \Sigma_0)$ can be estimated by optimising a data driven criterion or using the EM algorithm. In addition to the three main approaches to estimation outlined above, numerous alternatives have also been proposed. These include approximating the posterior density by piecewise linear functions (Kitagawa (1987)), using numerical integration methods (Fruhwirth-Schnatter (1994)), and particle filters Kitagawa (1996), none of which are use here.

Algorithm 2. *Kalman filter and smoother*
For iteration j let $\mathbf{b}_{i|T} = \mathbb{E}[\boldsymbol{\beta}_i|y_1,\ldots,y_T]$ and $V_{i|T} = \mathrm{Var}[\boldsymbol{\beta}_i|y_1,\ldots,y_T]$ be the estimate and variance of $\boldsymbol{\beta}_i$ given data up to y_T.

1. *Initialise the process by setting $\mathbf{b}_{0|0} = \mu_0$ and $V_{0|0} = \Sigma_0$.*

2. *For $i = 1,\ldots,n$ calculate the Kalman filter estimates $\mathbf{b}_{i|i}$ and variances $V_{i|i}$ using the equations*

$$\begin{aligned}
\mathbf{b}_{i|i-1} &= \Phi\mathbf{b}_{i-1|i-1}, \\
V_{i|i-1} &= \Phi V_{i-1|i-1}\Phi^T + \Sigma_\beta, \\
\mathbf{b}_{i|i} &= \mathbf{b}_{i|i-1} + K_i[\tilde{y}_i(\boldsymbol{\beta}^{(j)}) - \mathbf{x}_i^T\mathbf{b}_{i|i-1}], \\
V_{i|i} &= V_{i|i-1} - K_i\mathbf{x}_i^T V_{i|i-1}, \\
K_i &= V_{i|i-1}\mathbf{x}_i[\mathbf{x}_i^T V_{i|i-1}\mathbf{x}_i + W(\boldsymbol{\beta}^{(j)})]^{-1},
\end{aligned}$$

   in which $W(\boldsymbol{\beta}^{(j)})$ is the diagonal weight matrix with elements given by equation (2.5).

3. *For $i = n,\ldots,1$ calculate the Kalman smoother estimates $\mathbf{b}_{i|n}$ and variances $V_{i|n}$ using the equations*

$$\begin{aligned}
\mathbf{b}_{i-1|n} &= \mathbf{b}_{i-1|i-1} + B_i(\mathbf{b}_{i|n} - \mathbf{b}_{i|i-1}), \\
V_{i-1|n} &= V_{i-1|i-1} + B_i(V_{i|n} - V_{i|i-1})B_i^T, \\
B_i &= V_{i-1|i-1}\Phi^T V_{i|i-1}^{-1}.
\end{aligned}$$

4. *Set $\boldsymbol{\beta}_i^{(j+1)} = \mathbf{b}_{i|n}$ for all i.*

2.6.2 Temporal stochastic processes

Stochastic processes are commonly used to model trends and correlation in temporal data, and reviews are given by Cox and Miller (1968) and Chatfield (1996). In the air pollution literature autoregressive processes have been used for this purpose (see for example Schwartz et al. (1996)), including within the context of a dynamic generalised linear model (Chiogna and Gaetan (2002)). An autoregressive processes of order p (denoted AR(p)) is a discrete time processes $\{V_i\}_{i=1}^{\infty}$ which is given by

$$V_i = \theta_1 V_{i-1} + \ldots + \theta_p V_{i-p} + \epsilon_i, \tag{2.10}$$

where ϵ_i is a zero mean random error that is typically assumed to be Gaussian. The evolution of the process is controlled by the state parameters $\theta = (\theta_1,\ldots,\theta_p)$, which

determine its defining characteristics such as mean, variance and stationarity (see below). An autoregressive process is a special case of the wider class of autoregressive integrated moving average processes (ARIMA(p,d,q) Box and Jenkins (1976)) that have a general form given by

$$\phi(B)(1-B)^d V_i = \psi(B)\epsilon_i.$$

Here B denotes the backshift operator, meaning that $V_{i-1} = BV_i$, while ϕ and ψ are polynomials in B given by $\phi(B) = 1 - \theta_1 B - \ldots - \theta_p B^p$ and $\psi(B) = 1 + \beta_1 B + \ldots + \beta_q B^q$ respectively. The component $(1-B)^d$ allows the processes to be differenced, which is typically required if the data being modelled have a time-varying mean or variance. The autoregressive component of this model is specified by $\phi(B)V_i$, and (2.10) is obtained by setting $d = 0$ and $\psi(B) = 1$. An important subclass of temporal processes are those that exhibit time invariant characteristics, which are called stationary. A temporal process is strictly stationary if the joint distribution of $f(V_{i_1}, \ldots, V_{i_n})$ is the same as the joint distribution of $f(V_{i_1+\tau}, \ldots, V_{i_n+\tau})$ for all i and τ. However this criteria is overly restrictive and is rarely achieved in practice. As a result a weaker criteria called second order stationarity is more often used. A stochastic process is second order stationary if

- $\mathbb{E}[V_i] = \mu$,
- $\mathrm{Cov}[V_i, V_{i+\tau}] = \gamma(\tau)$ for all i,
- The mean and covariance are finite.

Strict stationarity forces the distribution of the temporal process to be constant over time, while second order only restricts the mean and covariance function in this way. Strict stationarity implies second order stationary but the reverse is not true unless the data are Gaussian. The stationarity of an ARIMA process depends on $\phi(B)$ and d, and if the latter is greater than zero the process is non-stationary. In contrast if $d = 0$ the process is stationary as long as the roots of $\phi(B)$ lie outside the unit circle. In particular, a first order autoregressive process is stationary if $|\theta_1| < 1$.

2.7 Varying coefficient models

Varying coefficient models (VCM, Hastie and Tibshirani (1993)) extend generalised linear, additive and dynamic models by incorporating interactions between pairs of covariates. The response data come from an exponential family distribution f, whose mean is related to covariates $\{x_j\}_{j=1}^q$ and $\{r_j\}_{j=1}^q$ by a known and invertible link function g. The first set of covariates have the same meaning as in previous sections, while their corresponding regression parameters are functions of the second set $\{r_j\}_{j=1}^q$. These additional covariates

are called effect modifiers, and the parameter functions can be simple parametric forms or flexible smooth functions. A varying coefficient model is given by

$$\begin{aligned} y_i &\sim f(y_i|\mu_i) \quad \text{for} \quad i = 1, \ldots, n, \\ g(\mu_i) &= \eta_i = \beta_0 + \sum_{j=1}^{q} x_{ij}\beta_j(r_{ij}), \end{aligned} \tag{2.11}$$

and a number of simpler models can be regained by particular specifications of x_{ij} and $\beta_j(r_{ij})$. These include:

- a generalised linear model if $\beta_j(r_{ij}) = \beta_j$, a constant function;
- a generalised additive model if $x_{ij} = 1$ and $\beta_j(r_{ij})$ is a non-parametric smooth function;
- a dynamic generalised linear model if $r_{ij} = i$ and $\beta_j(i) \sim \mathrm{N}(\phi\beta_j(i-1), \sigma^2)$.

In addition to these special cases setting $r_{ij} = i$ results in a time-varying coefficient model (TVCM), which is used in chapter five to investigate the possibility of temporal variation in the pollution-mortality relationship. A brief review of estimation for varying coefficient models is given below.

2.7.1 Estimation

Varying coefficient models are too general for a single estimation algorithm, and additional assumptions about $\beta_j(r_{ij})$ are required. If the model reduces to the simplified forms above estimation using Bayesian or likelihood methods is straightforward and has been described in previous sections. If $\beta_j(r_{ij})$ are additive parametric functions, that is $\beta_j(r_{ij}) = \mathbf{w}_i(r_{ij})^{\mathrm{T}}\boldsymbol{\theta}$ for known $\mathbf{w}_i(r_{ij})$ and unknown parameters $\boldsymbol{\theta}$, estimation is also straightforward because the model reduces to a generalised linear model. In contrast if $\beta_j(r_{ij})$ are smooth non-parametric functions estimation can be based on the penalised least squares criterion proposed by Hastie and Tibshirani (1993).

2.8 Measurement error models

Measurement error is a general term used to encompass situations where the observed data do not represent the quantity of interest exactly. It can occur in both response variables and covariates, but only the latter are discussed here because error in $\mathbf{y}$ is not investigated in this thesis. In chapter six the existence of measurement error in the ambient pollution data is investigated, and a brief review of measurement error models is given below. More comprehensive discussions are given by Fuller (1987) and Carroll et al. (2006), which focus

on linear and non-linear models respectively. Measurement error models are based on four quantities

- $\mathbf{y}_{n\times 1}$ - a vector of response variables;
- $\mathbf{x}_{n\times 1}$ - a vector of true unobserved exposures ;
- $\mathbf{w}_{n\times 1}$ - a vector of observed exposures that are mis-measured version of $\mathbf{x}$;
- $Z_{n\times q}$ - a matrix of q covariates that are measured exactly.

The joint likelihood of these quantities can be decomposed into

$$f(\mathbf{y}, \mathbf{x}, \mathbf{w}|Z) = f(\mathbf{y}|\mathbf{x}, \mathbf{w}, Z) f(\mathbf{x}, \mathbf{w}|Z), \tag{2.12}$$

where the covariates are conditioned on because they are fixed and known. The first element of (2.12) is an exposure-response model, which can be any of the regression models described in previous sections. However it is typically simplified to $f(\mathbf{y}|\mathbf{x}, Z)$ by assuming the measurement error is non-differential, that is $\mathbf{y}$ and $\mathbf{w}$ are conditionally independent given $\mathbf{x}$. The second element in (2.12) is a measurement error model, which represents the relationship between the unobserved exposure $\mathbf{x}$ and the measured surrogate $\mathbf{w}$. There are two types of measurement error model, called classical and Berkson, both of which are outlined below.

2.8.1 Classical measurement error

Classical measurement error models decompose $f(\mathbf{x}, \mathbf{w}|Z)$ into $f(\mathbf{w}|\mathbf{x}, Z) f(\mathbf{x}|Z)$, the first element of which is a conditional model for the measured surrogate $\mathbf{w}$ given the true (unobserved) exposure $\mathbf{x}$. Two common classical measurement error models are called additive and error calibration, and are given by (i) and (ii) below. In both cases $f(\mathbf{w}|\mathbf{x}, Z) = \prod_{i=1}^{n} f(w_i|x_i, \mathbf{z}_i)$, a product of independent models for each observation.

$$\begin{aligned} &\text{(i)} \quad w_i \sim \mathrm{N}(x_i, \sigma^2) \quad \text{for } i = 1, \ldots, n \\ &\text{(ii)} \quad w_i \sim \mathrm{N}(\theta_1 + \theta_x x_i + \mathbf{z}_i^{\mathrm{T}}\theta, \sigma^2) \quad \text{for } i = 1, \ldots, n \end{aligned}$$

In the simple additive formulation the observed surrogate is assumed to be correct on average (that is $\mathbb{E}[w_i|x_i] = x_i$), while in model (ii) the surrogate is biased. Both models specify an additive relationship between (w_i, x_i), and a further alternative is to use a multiplicative error model $w_i = x_i \epsilon_i$, where ϵ_i is a zero mean Gaussian error with variance σ^2. The remaining term $f(\mathbf{x}|Z)$ can be based on knowledge of the true exposure or represent prior ignorance, and in the latter case a common choice is $\prod_{i=1}^{n} \mathrm{N}(x_i|\mu, \tau^2)$, where τ^2 is large. In a Bayesian setting $f(\mathbf{x}|Z)$ acts as a prior for the unknown exposure $\mathbf{x}$.

2.8.2 Berkson measurement error

In contrast Berkson measurement error models set $f(\mathbf{x}, \mathbf{w}, Z) = f(\mathbf{x}|\mathbf{w}, Z)f(\mathbf{w}|Z)$, where the first term is a conditional model for the true exposure $\mathbf{x}$ given the measured surrogate $\mathbf{w}$. Two common Berkson models are called additive and regression calibration, and are given by (iii) and (iv) below. In common with the classical models $f(\mathbf{x}|\mathbf{w}, Z)$ is a product of independent distributions for each observation.

$$\begin{aligned} \text{(iii)} \quad x_i &\sim \mathrm{N}(w_i, \sigma^2) \quad \text{for } i = 1, \ldots, n \\ \text{(iv)} \quad x_i &\sim \mathrm{N}(\theta_1 + \theta_w w_i + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\theta}, \sigma^2) \quad \text{for } i = 1, \ldots, n \end{aligned}$$

In the simple additive model the true exposure is assumed to be equal to the surrogate on average (that is $\mathbb{E}[x_i|w_i] = w_i$), but this is not true for (iv). As with classical models a multiplicative alternative can be used, which is implemented using an additive model on the log scale. In the Berkson model $f(\mathbf{w}|Z)$ can be ignored because $\mathbf{w}$ is a known measurement, meaning that its distribution is redundant. The choice between classical and Berkson models will depend on the structure of the problem as well as the set of available data, and further details are given by Carroll et al. (2006).

The measurement error models described above can only be used if additional data are available, because $(\mathbf{y}, \mathbf{w}, Z)$ are not sufficient to estimate the measurement error process. Examples of such additional data include repeated measurements of $\mathbf{w}$, observed true exposures $\mathbf{x}$ for a subset of the n observations, or external data including observed values of $\mathbf{x}$ and $\mathbf{w}$. The identifiability of a proposed model may also depend on the assumptions made about the measurement error process, and there are two generic classes, functional and structural. Functional models are distribution invariant and specify minimal assumptions about the measurement error process. In contrast structural models (such as (i) to (iv)) are fully parametric, and specify probability distributions for $f(\mathbf{w}|\mathbf{x}, Z)$ or $f(\mathbf{x}|\mathbf{w}, Z)$. The choice between functional and structural models determines the method of estimation and inference that can be used, with structural models enabling likelihood and Bayesian methods to be applied. In contrast functional models do not specify a proper likelihood, and estimation is typically based on regression calibration and the SIMEX algorithm. A brief outline of estimation techniques for measurement error models is given below, while a more comprehensive treatment is given by Carroll et al. (2006).

2.8.3 Estimation

The measurement error models applied in chapter six are based on log normal distributions, so only structural estimation techniques such as likelihood and Bayesian methods are outlined here. The parameters in the model are collectively denoted by $\boldsymbol{\Omega}$, which can be

estimated by maximising the joint likelihood of $(\mathbf{y}, \mathbf{w})$ given by

$$f(\mathbf{y}, \mathbf{w}|Z, \mathbf{\Omega}) = \int_{\mathbf{x}} f(\mathbf{y}, \mathbf{x}, \mathbf{w}|Z, \mathbf{\Omega}) d\mathbf{x}. \tag{2.13}$$

As previously described $f(\mathbf{y}, \mathbf{x}, \mathbf{w}|Z, \mathbf{\Omega})$ can be factorised into classical or Berkson error models:

$$\begin{array}{lrcl} \text{Classical} & f(\mathbf{y}, \mathbf{x}, \mathbf{w}|Z, \mathbf{\Omega}) & = & f(\mathbf{y}|\mathbf{x}, Z, \omega_1) f(\mathbf{w}|\mathbf{x}, Z, \omega_2) f(\mathbf{x}|Z, \omega_3) \quad \text{or} \\ \text{Berkson} & f(\mathbf{y}, \mathbf{x}, \mathbf{w}|Z, \mathbf{\Omega}) & = & f(\mathbf{y}|\mathbf{x}, Z, \omega_1) f(\mathbf{x}|\mathbf{w}, Z, \omega_2) f(\mathbf{w}|Z, \omega_3), \end{array}$$

where $\mathbf{\Omega} = (\omega_1, \omega_2, \omega_3)$. However in the Berkson setting $\mathbf{w}$ is assumed to be a known covariate, meaning that (2.13) is replaced by

$$f(\mathbf{y}|Z, \mathbf{w}, \mathbf{\Omega}) = \int_{\mathbf{x}} f(\mathbf{y}|\mathbf{x}, Z, \omega_1) f(\mathbf{x}|\mathbf{w}, Z, \omega_2) d\mathbf{x}, \tag{2.14}$$

where $\mathbf{w}$ has been conditioned out of the joint likelihood. Likelihood methods estimate $\mathbf{\Omega}$ by maximising $f(\mathbf{y}, \mathbf{w}|Z, \mathbf{\Omega})$ or $f(\mathbf{y}|Z, \mathbf{\Omega})$, where ω_1 is of primary interest because it describes the relationship between the response $\mathbf{y}$ and the true exposure $\mathbf{x}$. Neither (2.13) nor (2.14) are typically available in closed form, and estimation can be achieved using a variety of techniques that account for the integral over $\mathbf{x}$. In simple problems where the likelihood can be computed or well approximated analytically $\mathbf{\Omega}$ can be estimated by iterative numerical methods. An example is the work of Schafer (2001), who uses the EM algorithm for estimation in semi-parametric structural models. In more complex problems monte-carlo techniques can be used to approximate $f(\mathbf{y}, \mathbf{w}|Z, \mathbf{\Omega})$ or $f(\mathbf{y}|Z, \mathbf{\Omega})$, an example of which is given by Geyer and Thompson (1992). Alternatively measurement error models can be viewed as missing data problems (where $\mathbf{x}$ is a missing covariate), and estimation methods can be borrowed from that literature (for a review see Little and Rubin (2002)).

Bayesian measurement error models comprise one of the likelihoods given by equations (2.13) or (2.14) and a prior $f(\mathbf{\Omega})$, the latter of which is a product of marginal and conditional distributions. Bayesian inference is based on the posterior distribution of $\mathbf{\Omega}$, which for classical and Berkson models is proportional to

$$\begin{array}{lrcl} \text{Classical} & f(\mathbf{\Omega}|\mathbf{y}, \mathbf{w}, Z) & \propto & f(\mathbf{\Omega}) \int_{\mathbf{x}} f(\mathbf{y}|\mathbf{x}, Z, \omega_1) f(\mathbf{w}|\mathbf{x}, Z, \omega_2) f(\mathbf{x}|Z, \omega_3) d\mathbf{x} \quad \text{or} \\ \text{Berkson} & f(\mathbf{\Omega}|\mathbf{y}, \mathbf{w}, Z) & \propto & f(\mathbf{\Omega}) \int_{\mathbf{x}} f(\mathbf{y}|\mathbf{x}, Z, \omega_1) f(\mathbf{x}|\mathbf{w}, Z, \omega_2) d\mathbf{x}. \end{array}$$

The posterior distribution is typically calculated using Markov chain monte carlo sim-

ulation, where $\mathbf{x}$ is treated as additional parameters to be estimated. Both Gibbs and Metropolis-Hastings algorithm have been used, and examples include conditionally independent models (Richardson and Gilks (1993)), non-linear regression models (Dellaportas and Smith (1993) and Berry et al. (2002)), and Gaussian mixture models (Richardson et al. (2002)).

2.8.4 Attenuation and bias

Covariate measurement error typically causes non-measurement error models to produce biased estimates of the regression parameters, and except for the simple linear model $y_i \sim \mathrm{N}(\beta_0 + \beta_x x_i, \sigma_\epsilon^2)$, the nature of this bias is largely unknown. In this simple case replacing $\mathbf{x}$ by $\mathbf{w}$ yields the model $y_i \sim \mathrm{N}(\beta_0^* + \beta_w w_i, \sigma_\epsilon^2)$, which estimates β_w instead of the relationship of interest β_x. As $\mathbf{x}$ is unknown β_x can only be estimated using measurement error methods, and a simple classical model is given by

$$\begin{aligned} y_i &\sim \mathrm{N}(\beta_0 + \beta_x x_i, \sigma_\epsilon^2), \\ w_i &\sim \mathrm{N}(x_i, \sigma_u^2), \\ x_i &\sim \mathrm{N}(\mu, \sigma_x^2). \end{aligned} \tag{2.15}$$

For this model it can be shown (for details see Fuller (1987)) that $\beta_w = \frac{\sigma_x^2}{\sigma_u^2 + \sigma_x^2}\beta_x$, so ignoring measurement error and naively replacing $\mathbf{x}$ with $\mathbf{w}$ results in a biased estimate of β_x. This bias is known as attenuation, and shrinks β_x towards zero by a factor of $\sigma_x^2/(\sigma_u^2 + \sigma_x^2)$. In addition, allowing for measurement error inflates the variance from $\mathrm{Var}[y_i|w_i] = \sigma_\epsilon^2$ using the simple linear model to $\mathrm{Var}[y_i|x_i] = \sigma_\epsilon^2 + \frac{\beta_x^2 \sigma_u^2 \sigma_x^2}{\sigma_x^2 + \sigma_u^2}$ if (2.15) is used. In contrast for the Berkson error model

$$\begin{aligned} y_i &\sim \mathrm{N}(\beta_0 + \beta_x x_i, \sigma_\epsilon^2), \\ x_i &\sim \mathrm{N}(w_i, \sigma_u^2), \end{aligned}$$

there is no attenuation, meaning that $\beta_w = \beta_x$. However in common with the classical model incorporating measurement error inflates the variance to $\mathrm{Var}[y_i|x_i] = \sigma_\epsilon^2 + \beta_x^2 \sigma_u^2$. Although the effects of measurement error are well known for the Gaussian linear model above, the corresponding effects for more complex linear and non-linear models have no exact analytical representation.

2.9 Spatio-temporal models

Spatio-temporal models extend those described in the previous six sections by incorporating the spatial structure in the data. Spatial data can be observed at any point in a continuous region (geostatistical) or at a set of discrete points or sub-regions (lattice data), and while air pollution data are geostatistical, mortality counts are an example of lattice data. The spatio-temporal regression models used in chapter six only apply to the pollution data, so a brief review of geostatistical spatial models is given below. For more general reviews of spatial and spatio-temporal models, see Banerjee et al. (2003) and Schabenberger and Gotway (2005). Letting $w_i(\mathbf{s})$ denote a spatio-temporal observation on day i at spatial location $\mathbf{s} = [s_1, s_2]$, the spatio-temporal regression model used both in chapter six and the wider air pollution literature has the general form

$$\begin{aligned} g(w_i(\mathbf{s}_l)) &\sim \mathrm{N}(g(x_i(\mathbf{s}_l)), \sigma_\epsilon^2 I) \quad \text{for } i = 1, \ldots, n \;\; l = 1, \ldots, k, \\ g(x_i(\mathbf{s}_l)) &= \mu_i(\mathbf{s}_l|\delta) + V_i(\mathbf{s}_l|\phi). \end{aligned} \tag{2.16}$$

The function g is a known transformation such as natural log or square root, which is applied to ensure the data are close to Gaussian. The first line of equation (2.16) is a measurement error model, specifying that each observation is equal to its true (unobserved) value $g(x_i(\mathbf{s}_l))$ plus classical measurement error. At the second stage this unobserved value is represented by a mean model $\mu_i(\mathbf{s}_l|\alpha)$ and a spatio-temporal stochastic process $V_i(\mathbf{s}_l|\phi)$, the latter of which models temporal and spatial correlation between the observations. The mean function can be a spatio-temporal extension to any of the models discussed in previous sections, while the spatio-temporal process is discussed in more detail below. The model in equation (2.16) is very general, and for estimation to be implemented further assumptions are required. Likelihood based methods use maximum likelihood and generalised least squares algorithms, while Bayesian methods incorporate MCMC simulation. A Bayesian simulation algorithm for a particular specification of this model is outlined in chapter six, while more general reviews of estimation are presented by Gotway and Stroup (1997) and Wikle et al. (1998).

2.9.1 Spatio-temporal processes

A spatio-temporal stochastic process is a collection of random variables

$$\mathbf{V} = \{V_i(\mathbf{s})|\mathbf{s} \in D \subset \mathbb{R}^2, i \in 1, 2, \ldots\},$$

which sit in a continuous spatial domain D. The processes used in chapter six are Gaussian with zero mean, because they allow correlation to be specified through the covariance matrix. The values in this matrix are typically specified by a parametric covariance function,

and a common simplification is that it is separable in time and space. This separability has two forms multiplicative or additive, which occur if the covariance function can be written as

$$\begin{aligned} \text{Cov}[V_i(\mathbf{s}), V_{i+\tau}(\mathbf{s}+\mathbf{h})] &= \gamma_s(\mathbf{s}, \mathbf{s}+\mathbf{h})\gamma_i(i, i+\tau) \quad \text{or} \\ \text{Cov}[V_i(\mathbf{s}), V_{i+\tau}(\mathbf{s}+\mathbf{h})] &= \gamma_s(\mathbf{s}, \mathbf{s}+\mathbf{h}) + \gamma_i(i, i+\tau), \end{aligned}$$

the product or sum of spatial $\gamma_s(\mathbf{s}, \mathbf{s}+\mathbf{h})$ and temporal $\gamma_i(i, i+\tau)$ covariance functions. Separable covariance functions are commonly used in the air pollution literature (see for example Shaddick and Wakefield (2002) and Zhu et al. (2003)) because they are computationally cheaper to implement than non-separable alternatives. However separable functions are restrictive because they do not allow space-time interactions, meaning that the spatial covariance structure at different time points is identical. The covariance functions used in chapter six are separable for the reasons outlined above, but examples of non-separable covariance functions are given by Cressie and Huang (1999) and Gneiting (2002). As temporal stochastic processes were discussed in section 2.6.2, the remainder of this section focuses on spatial processes.

A geostatistical spatial process is also known as a random field, and is a collection of correlated random variables $\mathbf{V} = \{V(\mathbf{s}) | \mathbf{s} \in D \subset \mathbb{R}^d\}$. Two subclasses are those that are stationary and isotropic, which force the process to exhibit spatially invariant characteristics. A spatial random field is strictly stationary if its distribution is invariant to a change in the coordinates, that is if $f(V(\mathbf{s}_1), \ldots, V(\mathbf{s}_k))$ is the same as $f(V(\mathbf{s}_1+\mathbf{h}), \ldots, V(\mathbf{s}_k+\mathbf{h}))$ for any k and $\mathbf{h} \in \mathbb{R}^d$. However in common with temporal processes this is overly restrictive, and a weaker criteria is second order stationary. A spatial process is second order stationary if

- $\mathbb{E}[V(\mathbf{s}_i)] = \mu$,
- $\text{Cov}[V(\mathbf{s}_i), V(\mathbf{s}_i+\mathbf{h})] = \gamma(\mathbf{h})$ for all i and $\mathbf{h} \in \mathbb{R}^d$,
- the mean and covariance are both finite.

Stationarity forces the covariance between $(V(\mathbf{s}), V(\mathbf{s}+\mathbf{h})$ to only depend on the separation vector $\mathbf{h}$, meaning that the covariance of two observations does not depend on the original spatial location $\mathbf{s}$. A further simplification is isotropy, which occurs if $\text{Cov}[V(\mathbf{s}_i), V(\mathbf{s}_i+\mathbf{h})] = \gamma(||\mathbf{h}||)$ where $||.||$ denotes the Euclidean norm. A common class of stationary and isotropic covariance functions is the Matérn class (Matérn (1986)), whose general form is given by

$$\text{Cov}[V(\mathbf{s}_i), V(\mathbf{s}_i+\mathbf{h})] = \sigma^2 \frac{1}{\Gamma(\nu)} \left(\frac{\theta||\mathbf{h}||}{2}\right) 2K_\nu(\theta||\mathbf{h}||) \quad \text{for } \nu, \theta, \sigma^2 > 0.$$

In the above equation $K_\nu(\theta||\mathbf{h}||)$ is a modified Bessel function of the second kind, and $(\nu,\ \theta,\ \sigma^2)$ are parameters that control the strength of the correlation. Two special cases are the Gaussian and exponential correlation models, which occur when $\nu = \infty$ and $\nu = 1/2$, resulting in

$$\begin{aligned} \text{Gaussian} \quad & \gamma(||\mathbf{h}||) = \sigma^2 \exp(-\theta||\mathbf{h}||^2) \quad \text{and} \\ \text{exponential} \quad & \gamma(||\mathbf{h}||) = \sigma^2 \exp(-\theta||\mathbf{h}||) \end{aligned}$$

respectively. Although the majority of studies that model the underlying pollution surface use stationary and isotropic covariance functions (see for example Shaddick and Wakefield (2002)), non-stationary and non-isotropic alternatives have been used, notably by Fuentes (2002) and Fuentes and Raftery (2005).

2.10 Model selection and adequacy

The aim of regression modelling is to produce a good description of $\mathbf{y}$ in terms of its covariate risk factors, which may be achieved by a number of candidate models. This set of models may differ in a number of respects including the probability model for $\mathbf{y}$; the link function g; the matrix of covariates X; the form in which each covariate enters the model (for example as a linear effect or an unknown smooth function); and the prior distribution $f(\theta)$ (Bayesian analyses only). Changing any of these factors may affect the substantive conclusions drawn from the analysis, and a number of methods for facilitating model choice are available. These methods address two related problems:

1. Given a selection of candidate models that appear to describe the data well, which model or models should be used for inference?

2. Once a 'final' model has been chosen, how can it's adequacy as a description of the data be assessed?

Both of these are addressed below, although the distinction between selection and adequacy is somewhat arbitrary because most of the methods can be used for both purposes. However the techniques described in this section should not be used in isolation, but rather in conjunction with personal judgement and experience.

2.10.1 Model selection

The model selection techniques outlined below are split into two types, those that measure how well a model fits a set of data and those that compare the performance of two competing models.

Overall measures of model fit

Schematically a regression model partitions the variation in a data set into

$$\text{Data} \;=\; \text{Fit} \;+\; \text{Residual},$$

in which 'Fit' represents the variation explained by the model, while 'Residual' represents the unexplained variation. For a data set $\mathbf{y}$ a regression model $\mathcal{M}$ lies between two extremes, the null and saturated models. The null model only contains an intercept term, effectively assigning all the variation in the data to the residual component. In contrast the saturated model has an equal number of data points and parameters, effectively assigning all the variation to the fit component. Consequently the adequacy of a model can be described by the proportion of variation it assigns to the fit component, with better performing models having smaller residual terms. This can be measured by the log-likelihood $\mathcal{L}(\hat{\theta}|\mathbf{y}, \mathcal{M})$, with larger values indicating a better fit to the data. The vector of parameters (θ) can be estimated by maximum likelihood methods, and $\mathcal{L}(\hat{\theta}|\mathbf{y}, \mathcal{M})$ lies between the corresponding values for the null and saturated models (denoted by $\mathcal{L}(\hat{\theta}_{null}|\mathbf{y})$ and $\mathcal{L}(\hat{\theta}_{max}|\mathbf{y})$ respectively). For the saturated model the fitted values are equal to the observed data, meaning that $\mathcal{L}(\hat{\theta}_{max}|\mathbf{y})$ is the maximum value that can be obtained for these data. The adequacy of a model is typically assessed by

$$D(\hat{\theta}|\mathbf{y}, \mathcal{M}) = 2[\mathcal{L}(\hat{\theta}_{max}|\mathbf{y}) - \mathcal{L}(\hat{\theta}|\mathbf{y}, \mathcal{M})],$$

which is called the deviance (Nelder and Wedderburn (1972)) and represents the difference between $\mathcal{M}$ and the saturated model. Therefore a set of candidate models can be compared by calculating their respective deviances, with smaller values suggesting a better fit to the data. The deviance can also be used to assess the adequacy of a single model rather than as a comparative tool. If model $\mathcal{M}$ is an adequate description of the data

$$D(\hat{\theta}|\mathbf{y}, \mathcal{M}) \overset{\text{approx}}{\sim} \chi^2_{n-q},$$

where q is the number of parameters in model $\mathcal{M}$. The approximation improves asymptotically as the number of data points increases, and large deviances (typically values that occur less than 5% of the time under a χ^2_{n-q} distribution) suggest that $\mathcal{M}$ is not an adequate description of the data. However the deviance does not take into account the number of parameters in a model, so the addition of an extra covariate will lower the deviance (suggesting a better fit to the data) regardless of whether it is causally related to $\mathbf{y}$. As a result model selection is typically based on alternative criteria, that account for the number of parameters in a model. A number of these criteria still measure closeness to the data using the deviance, but include a second term that penalises models with an excessive number of parameters. Two examples are Akaike's information criterion (AIC Akaike (1973))

$$\text{AIC} = \frac{D(\hat{\theta}|\mathbf{y}, \mathcal{M})}{n} + \frac{2q}{n},$$

and generalised cross validation (GCV)

$$\text{GCV} = \frac{\frac{1}{n}D(\hat{\theta}|\mathbf{y}, \mathcal{M})}{(1 - \frac{1}{n}q)^2},$$

where in both cases q is the effective number of parameters in model $\mathcal{M}$, and small values suggest a good fit to the data. Other model selection criteria include Bayesian information criteria (BIC), Mallow's C_p and the PRESS criterion, but AIC and GCV are listed here because they are used in chapters four and five. However determining the effective number of parameters in a model may not be straightforward, especially if it has a hierarchical structure such as a dynamic generalised linear model. The parameters in this model are restricted by an autoregressive constraint, meaning that the effective number of parameters is much less than the total used in the model. As a result estimating this quantity is not straightforward, although it is typically approximated by the trace of the generalised hat matrix (Ruppert et al. (2003)).

The criteria described above are used for likelihood based inference, and a Bayesian alternative is the Deviance Information Criterion (DIC Spiegelalter et al. (2002)). The Bayesian deviance for model $\mathcal{M}$ is given by $D_B(\boldsymbol{\theta}|\mathbf{y}, \mathcal{M}) = -2\mathcal{L}(\boldsymbol{\theta}|\mathbf{y}, \mathcal{M})$, which differs from the likelihood definition by the constant $2\mathcal{L}(\boldsymbol{\theta}_{max}|\mathbf{y})$. However there are no natural point estimates in a Bayesian setting, and $D_B(\boldsymbol{\theta}|\mathbf{y}, \mathcal{M})$ can be estimated in two ways.

(i) Set $D_{\tilde{\boldsymbol{\theta}}}(\mathbf{y}|\mathcal{M}) = D_B(\tilde{\boldsymbol{\theta}}|\mathbf{y}, \mathcal{M})$ where $\tilde{\boldsymbol{\theta}}$ is the posterior mean of $\boldsymbol{\theta}$.

(ii) Average the deviance over the posterior distribution of $\boldsymbol{\theta}$, that is set $D_{\text{av}}(\mathbf{y}|\mathcal{M}) = \mathbb{E}[D(\boldsymbol{\theta}|\mathbf{y}, \mathcal{M})|\mathbf{y}]$. This can be estimated by simulation as $\hat{D}_{\text{av}}(\mathbf{y}|\mathcal{M}) = (1/L)\sum_{j=1}^{L} D(\boldsymbol{\theta}^{(j)}|\mathbf{y}, \mathcal{M})$, where $\{\boldsymbol{\theta}^{(j)}\}_{j=1}^{L}$ are L samples from the posterior distribution of $\boldsymbol{\theta}$.

Here $D_{\tilde{\boldsymbol{\theta}}}(\mathbf{y}|\mathcal{M})$ is always smaller than $\hat{D}_{\text{av}}(\mathbf{y}|\mathcal{M})$ because $\tilde{\boldsymbol{\theta}}$ provides a better fit to the data then averaging over the posterior distribution. As a result $P_D(\mathcal{M}) = \hat{D}_{\text{av}}(\mathbf{y}|\mathcal{M}) - D_{\tilde{\boldsymbol{\theta}}}(\mathbf{y}|\mathcal{M})$ represents the effect of model fitting, and is a measure of the effective number of parameters in a Bayesian model. The deviance information criteria is given by

$$\text{DIC}(\mathcal{M}) = D_{\tilde{\boldsymbol{\theta}}}(\mathbf{y}|\mathcal{M}) + 2P_D(\mathcal{M}),$$

in which the first term measures model adequacy. The second term penalises models with an excessive number of parameters, and in common with the criteria outlined above better performing models have lower $\text{DIC}(\mathcal{M})$. Comprehensive reviews of model selection

techniques are given by Hastie and Tibshirani (1990), Miller (1990), Fahrmeir and Tutz (2001) and Ruppert et al. (2003).

Comparing two specific models

The model selection techniques described above are based on objective criteria, and an alternative strategy is to directly compare two similar models by means of a hypothesis test. For example the explanatory power of a covariate w can be assessed by carrying out a χ^2 test on two nested models. Let $\mathcal{M}_0$ be a candidate model with q free parameters, and consider an extension $\mathcal{M}_1 = \mathcal{M}_0 + w$ which has p free parameters where $p > q$. The statistical significance of w can be tested by calculating the test statistic

$$T = \frac{D(\hat{\theta}|\mathbf{y}, \mathcal{M}_0) - D(\hat{\theta}|\mathbf{y}, \mathcal{M}_1)}{a(\phi)} \overset{\text{approx}}{\sim} \chi^2_{p-q},$$

which has an approximate χ^2_{p-q} distribution if $\mathcal{M}_0$ is an adequate description of the data. This test is approximate for a general exponential family distribution but has an exact $F_{p-q,n-p}$ distribution if $\mathbf{y}$ is a multivariate Gaussian. This test also forms the basis of stepwise regression, an automatic variable selection algorithm for regression models. The algorithm is initialised at a starting model, which can be either of the null or saturated models. At each step the algorithm carries out the hypothesis test described above, to determine whether an additional variable should be added (a forward step) or removed (a backward step) from the current model. The algorithm terminates when the model reaches a steady state in which no variables can be added or removed. Stepwise regression is a fully algorithmic procedure which does not incorporate human judgement, and should therefore be used with caution. The starting model may affect the final model, while all possible combinations of covariates are not compared. Further details of stepwise regression and other model selection tests are given by Miller (1990).

Bayes factors (Kass and Raftery (1995)) are a Bayesian alternative to hypothesis tests, which compare a series of nested or non-nested models $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_r$. The models are compared via their posterior model probabilities $f(\mathcal{M}_k|\mathbf{y})$, which indicate the likelihood of the data originating from each model. These probabilities are based on the prior $f(\mathcal{M}_k)$ and the data likelihood $f(\mathbf{y}|\mathcal{M}_k)$, meaning that posterior model probabilities are given by

$$f(\mathcal{M}_k|\mathbf{y}) = \frac{f(\mathbf{y}|\mathcal{M}_k)f(\mathcal{M}_k)}{\sum_{k=1}^{r} f(\mathbf{y}|\mathcal{M}_k)f(\mathcal{M}_k)} \quad \text{for} \quad k = 0, \ldots, r.$$

However Bayes factors make the unrealistic assumption that only r candidate models exist, and are further complicated by the choice of prior which may not be straightforward.

2.10.2 Model adequacy

Once a final model has been chosen its adequacy as a description of the data should be assessed. A broad range of model checking methods are available, including overall measures of model adequacy, sensitivity analysis, residual based methods and posterior predictive checking, all of which are outlined below.

Overall measures of model adequacy

Numerous measures of model adequacy have been proposed, including the deviance-χ^2 test described in the previous section, and Pearson's goodness of fit test. The latter of these measures the distance between $\mathbf{y}$ and the fitted values from the model, and the test statistic is given by

$$T = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\text{Var}[y_i]} \overset{\text{approx}}{\sim} \chi^2_{n-q},$$

where $\hat{\mu}_i$ denotes a fitted value from the model. If the model is adequate T has an approximate χ^2_{n-q} distribution, where q is the effective number of parameters in the model. Further details of these and other tests of model adequacy are given by Fahrmeir and Tutz (2001) and Dobson (1990).

Sensitivity analysis

Sensitivity analysis applies a set of candidate models to the data to determine whether the choice of model affects the results. These models typically differ in a single aspect such as their prior, and if the results are robust to small changes in model specification conclusions can be based on any of the candidate models. However if the candidate models produce significantly different results they should be treated with caution, and the cause of these differences should be investigated. In a Bayesian setting sensitivity analysis is commonly used to assess the impact of the priors, and is carried out here in chapters four to six.

Residual based methods

Residual variation describes the difference between $\mathbf{y}$ and the fitted model $\mathcal{M}$, and if the latter is a good description of the data, it will resemble independent random fluctuations that contain no structure or correlation. This variation can be measured by a number of metrics, a review of which is given by Fahrmeir and Tutz (2001). The simplest is the raw residual $\tilde{r}_i = y_i - \hat{\mu}_i$, which measures the absolute difference between the data and the model. However these residuals are not ideal because each $\tilde{r}_i$ may have a different variance, meaning that outliers are difficult to detect. As a result standardised residuals

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{Var}[y_i]}} \overset{\text{approx}}{\sim} \text{N}(0,1) \quad \text{for } i = 1, \ldots, n,$$

are typically used for model assessment, and can be compared against ± 1.96 to identify possible outliers. In addition they can be used to check the distributional assumption for $\mathbf{y}$, by plotting the ordered $\{r_i\}_{i=1}^n$ against quantiles of the standard Gaussian distribution. A straight line suggests the distributional assumption is adequate, while curves suggest that a heavier or lighter tailed distribution is appropriate. Furthermore any structure in the residuals can be examined graphically, by plotting them against a range of covariates such as observation number. The presence of un-modelled correlation in $\mathbf{y}$ can also be investigated using the residuals, by calculating the sample autocorrelation and partial autocorrelation functions of $\{r_i\}_{i=1}^n$ (for details see Box and Jenkins (1976)). In the likelihood paradigm the residuals are based on the maximum likelihood estimate $\hat{\theta}$, whereas for a Bayesian analysis the posterior mean, median, or mode could all be used. Instead r_i can be summarised by a Bayesian residual distribution

$$f(r_i|\mathbf{y}) = \int_{\theta} f(r_i, \theta|\mathbf{y}) d\theta,$$

which averages over the posterior uncertainty in θ removing the need for a specific estimate. Consequently the likelihood residuals based on $\hat{\theta}$ can be thought of as an approximation to this residual distribution. It can be estimated by simulation as

$$r_i^{(j)} = \frac{y_i - \mu_i^{(j)}}{\sqrt{\text{Var}[y_i]^{(j)}}} \quad \text{for } j = 1, \ldots, L,$$

where $(\mu_i^{(j)}, \text{Var}[y_i]^{(j)})$ are based on the jth sample from $f(\theta|\mathbf{y})$. Each $r_i^{(j)}$ is called a realised residual and further details of Bayesian residuals are given by Pettit (1986) and Gelman et al. (2003).

Posterior predictive checking

Posterior predictive checking (Rubin (1984) and Gelman et al. (1996)) is a Bayesian tool for assessing model adequacy with respect to a particular facet of the data. If the model describes this facet of the data well then replicated data generated by the model should be similar to $\mathbf{y}$. That is $\mathbf{y}$ should be likely under its posterior predictive distribution

$$f(\mathbf{y}^{rep}|\mathbf{y}) = \int_{\theta} f(\mathbf{y}^{rep}|\theta) f(\theta|\mathbf{y}) d\theta,$$

where $\mathbf{y}^{rep}$ denotes data generated from the model. The posterior predictive distribution can be approximated by simulation, sampling θ from its posterior distribution and $\mathbf{y}^{rep}$ from $f(\mathbf{y}|\theta)$ given the sampled value of θ. Using the posterior predictive distribution model adequacy can be assessed by a test statistic or discrepancy measure $T(\mathbf{y}, \theta)$, which is a scalar summary of the data and parameters. The likelihood of $\mathbf{y}$ originating from the chosen model is assessed by its posterior predictive p-value

$$P_B = \mathbb{P}[T(\mathbf{y}^{rep}, \boldsymbol{\theta}) \geq T(\mathbf{y}, \boldsymbol{\theta})|\mathbf{y}],$$

which measures the extremity of the observed test quantity when compared with its posterior predictive distribution. The posterior predictive p-value can be calculated from L samples $\{\boldsymbol{\theta}^{(j)}, \mathbf{y}^{rep(j)}\}_{j=1}^{L}$ as the proportion for which $T(\mathbf{y}^{rep(j)}, \boldsymbol{\theta}^{(j)}) \geq T(\mathbf{y}, \boldsymbol{\theta}^{(j)})$. A precursor to posterior predictive checking was prior predictive checking (Box (1980)), which compares $\mathbf{y}$ to

$$f(\mathbf{y}^{rep}) = \int_{\boldsymbol{\theta}} f(\mathbf{y}^{rep}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

the marginal distribution of replicated data. This is called the prior predictive distribution, and can also be estimated by simulation sampling from $f(\boldsymbol{\theta})$ followed by $f(\mathbf{y}^{rep}|\boldsymbol{\theta})$. In particular the prior predictive distribution has no dependence on the observed data, and samples of $\boldsymbol{\theta}$ are drawn from the prior rather than the posterior. As a result prior predictive checking is sensitive to the choice of prior, and is undefined if an improper prior is used which limits it's utility especially in complex models.

Chapter 3

Air pollution and mortality studies

The association between short-term exposure to air pollution and mortality (or morbidity) is typically estimated from ecological data using Poisson linear or additive models. These data typically comprise area level summaries of mortality or morbidity, ambient pollution levels at fixed locations and meteorological covariates, all of which are routinely collected for other purposes. These are generally the only type of pollution and mortality data that are available, and their ecological nature present researchers with a number of statistical challenges. Examples include over-dispersion, temporal correlation, unmeasured confounding and misclassification of exposure, all of which need to be addressed when producing a realistic statistical model. In addition the association of interest is typically small, making estimation difficult and realistic models even more important. However as models become more realistic they may increase in complexity, requiring more computational power and time to estimate the parameters. Therefore the choice of statistical model results in a tradeoff between simple models that are computationally feasible and easy to interpret, and more complex alternatives which make less unrealistic assumptions about the data.

In this chapter I discuss the standard approaches to modelling these data and review the extensions proposed in the related literature. Similar reviews have recently been presented by Dominici et al. (2003) and Pope and Dockery (2006), who discuss the effects of long-term as well as short-term exposure to pollution. This chapter provides the context and motivation for the developments proposed in chapters four to six, and the first two sections critique the standard approaches to modelling air pollution and mortality data. Section 3.1 outlines the set of health, pollution and covariate data that are routinely used for this purpose, while section 3.2 reviews the standard statistical models. The next two sections describe components of the standard model in more detail. Section 3.3 focuses on the choices involved in specifying the air pollution component, while section 3.4 examines the choice and role of the covariates. The remainder of the chapter focuses on the limitations of air pollution and health studies, both in terms of the statistical models used and the

set of available data. Sections 3.5 and 3.6 describe the related problems of over-dispersion and temporal correlation which commonly afflict the health data. The last four sections discuss the conclusions that can be drawn from air pollution and health studies, with particular emphasis on extensions that provide stronger evidence of a pollution-mortality association. Section 3.7 outlines the limitations of single city studies and reviews the array of multi-city and meta analytic extensions. Section 3.8 concentrates on extensions to individual rather than ecological regression models, and outlines the strength of association that can be estimated from each type. Section 3.9 focuses on how pollution exposure is estimated, describing the standard approach and less restrictive alternatives. Finally section 3.10 examines the mortality displacement hypothesis, which affects the public health significance of air pollution and health studies.

3.1 Description of the data

The time series studies underpinning this thesis are based on daily ecological data relating to a geographical region $\mathcal{R}$ for n consecutive days. This region can be an extended urban area or a larger legislative district, and the analyses presented in chapters four to six are based on data from Greater London, Cleveland, Detroit, Minneapolis and Pittsburgh. These data comprise population based measures of mortality or morbidity, ambient pollution levels and area level covariates, all of which are described below.

3.1.1 Health data

Mortality or morbidity data are only available as aggregated daily counts, that are drawn from the population living within the geographical region $\mathcal{R}$. These data are denoted by $\mathbf{y} = (y_1, \ldots, y_n)_{n\times 1}$, where y_i represents the number of mortality or morbidity events that occur on day i. They are collected from hospital records and death registries, and for confidentiality reasons are not available at the individual level. All mortality events are classified by cause of death using the international classification of disease, which includes the data used in this thesis and that in the wider air pollution literature. The majority of these data are classified by the ninth revision (ICD-9 World Health Organisation (1975)), which covers the period from 1977 until 1999. A variety of mortality classifications have been used in such studies, the most general of which is total non-accidental mortality (ICD-9 codes <800, see for example Schwartz (1993)) which typically includes a large number of daily deaths. However it also includes a significant proportion of deaths unrelated to pollution exposure, which may cause the pollution-mortality association to be biased. Consequently cause specific classes such as mortality due to respiratory (ICD-9 codes 460-519, see for example Dab et al. (1996)) or cardiovascular (ICD-9 codes 390-459, see for example Vedal et al. (2003)) illness may be preferable, because they are more likely to be related to the possible effects of air pollution. However such a class may not

contain enough mortality events to permit accurate estimation of a pollution association. Numerous studies have also analysed mortality data relating to specific age groups such as the elderly (Dominici et al. (2000)) or children (Lin et al. (2002)), because this frail subpopulation are more likely to be susceptible to air pollution than the general population. In addition to mortality the association between air pollution exposure and morbidity has also been investigated, with positive associations found for asthma (see for example Yu et al. (2000)) and respiratory or circulatory illness (see for example Gwynn et al. (2000)). In chapters four to six both all cause (total non-accidental) and cause specific (respiratory) mortality data are analysed.

3.1.2 Air pollution data

The term 'air pollution' refers to a complex mixture of individual pollutants, a number of which are routinely measured. These measurements are typically made by k fixed site monitors, which are located at spatial co-ordinates $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_k\}$ throughout the region under study $\mathcal{R}$. These monitors measure continuously throughout the day and a daily average is calculated at each location. The daily averages relating to the period of study are collected in an $n \times k$ matrix $W(\mathcal{S})$, where the ilth element $w_i(\mathbf{s}_l)$ represents the average ambient measurement on day i at location $\mathbf{s}_l$. Numerous pollutants are routinely measured in this way, including particulate matter, sulphur dioxide, nitrogen dioxide, carbon monoxide, and ozone. The association between each of these pollutants and mortality (or morbidity) has been thoroughly investigated, although a sizeable proportion of recent studies relate to particulate matter, which is an '*air suspended mixture of solid and liquid particles that vary in size, composition, origins and effects*' (Dockery and Pope (1994)). Particulate matter is routinely measured by a number of different metrics including PM_{10} (which consists of particles that are less than $10\mu g/m^3$ in diameter), $PM_{2.5}$ (less than $2.5\mu g/m^3$ in diameter), total suspended particles (TSP), coefficient of haze (CoH) and Black smoke (BS). The first two of these are the subject of much recent research (see for example Laden et al. (2000) and Lin et al. (2002)), because it is hypothesised that fine particles are the most harmful to human health as they are able to travel deeper into the lungs. As a result PM_{10} is the main pollutant examined in chapters four to six, although carbon monoxide (CO), nitrogen dioxide (NO_2) and ozone (O_3) are also analysed in chapter six for comparative purposes.

3.1.3 Covariate data

Time series studies regress the health data against air pollution levels and a matrix of q covariates $Z = (\mathbf{z}_1^{\mathrm{T}}, \ldots, \mathbf{z}_n^{\mathrm{T}})_{n \times q}$, where $\mathbf{z}_i^{\mathrm{T}} = (z_{i1}, \ldots, z_{iq})_{q \times 1}$ denote the realisations for day i. These covariates model external risk factors that affect the daily health series, which typically induce long-term trends, seasonal variation, over-dispersion and temporal correlation into these data. The covariates are a crucial part of any air pollution study, because if

the influence of these factors is not adequately removed the estimated pollution-mortality association may be biased (confounded). The covariates typically used in these studies include meteorological conditions and artificial variables, the former of which are generally measured at airports. Examples of meteorological conditions include temperature, dew-point temperature, humidity, wind speed, number of sunshine hours, and daily rainfall, while artificial variables include functions of calendar time ($\{1, 2, \ldots, n\}$), existence of an influenza epidemic and indicator variables for 'day of the week'. These covariates are described in greater detail in section 3.4, and are used in chapters four to six to model both the pollution and mortality data.

3.2 Standard regression model

The health data described in the previous section are ecological, in the sense that they relate to the population as a whole rather than to individuals living within $\mathcal{R}$. As a result only a group level association between air pollution exposure and mortality or morbidity can be estimated, rather than being able to ascertain a causal exposure-response relationship. This group level association is typically estimated using generalised linear or additive models, in which the health data take the form of counts and are assumed to arise from a Poisson distribution. These data are regressed against air pollution levels and q covariates, the former being calculated from the $n \times k$ pollution matrix $W(\mathcal{S})$. These values are typically averaged over the k monitoring sites on each day, meaning that pollution exposure is estimated by $\mathbf{w} = (w_1, \ldots, w_n)_{n\times 1}$, where $w_i = \frac{1}{k}\sum_{l=1}^{k} w_i(\mathbf{s}_l)$. The model adopted in the majority of studies has the general form

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= w_i\gamma + \sum_{j=1}^{c_1} z_{ij}\alpha_j + \sum_{j=c_1+1}^{q} f_j(z_{ij}), \end{aligned} \tag{3.1}$$

where γ represents the group level association between air pollution exposure and mortality or morbidity. The first c_1 covariates enter the model as linear terms, while the remainder are incorporated as smooth functions f_j which allow a subset of the covariates to have non-linear effects. A wide range of covariates have been used in previous air pollution studies, including smooth functions of calendar time, trigonometric functions, meteorological conditions, an intercept term, and indicator variables for day of the week and influenza epidemics. The smooth functions have been estimated using parametric regression splines (Daniels et al. (2004)) and non-parametric smoothing splines (Dominici et al. (2000)), and while the former leads to a linear representation the latter results in an additive model. If regression splines are used the model simplifies to

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= w_i\gamma + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\alpha}, \end{aligned} \tag{3.2}$$

where $\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\alpha}$ may include the parametric regression splines. Estimation in air pollution studies is typically implemented within a frequentist setting using maximum likelihood methods, but with improvements in computational power Bayesian methods are becoming increasingly popular. If a Bayesian approach is adopted model (3.2) is generally used, because parametric functions are easier to implement in this setting than non-parametric alternatives. Consequently the standard Bayesian model is given by equation (3.2) and a non-informative Gaussian prior for $\boldsymbol{\beta} = (\gamma, \boldsymbol{\alpha})$. The estimated association between air pollution and health is typically transformed to the relative risk scale, which measures the impact on the population's health if the pollution level increased. For PM_{10} this risk is normally calculated for an increase of $10\mu g/m^3$, which for models (3.1) or (3.2) is given by

$$\text{RR}(\gamma) = \frac{\mu_i^{+10}}{\mu_i} = \exp(10 \times \gamma).$$

Here μ_i^{+10} is the expected number of deaths on day i if the PM_{10} level had increased by $10\mu g/m^3$. In this thesis a Bayesian approach is adopted throughout, and the smooth functions are represented by parametric regression splines. This is because they reduce to a linear combination of known covariates, making MCMC estimation straightforward to implement using a Metropolis-Hastings random walk proposal. In contrast non-parametric alternatives require more complex algorithms (see for example Hastie and Tibshirani (2000)), which would be computationally prohibitive to combine with the extensions proposed in chapters four to six. As a result the models proposed in these chapters are based on (3.2) rather than (3.1), both for the Bayesian implementations and the likelihood based comparisons. The next two sections describe the air pollution and covariate components of the standard regression models in greater detail.

3.3 Air pollution specification

In this section the air pollution component is described in greater detail, with particular focus on the choices of lag, pollutant and relationship with health.

3.3.1 Pollutant

As noted in section 3.1 researchers have found positive associations between mortality (or morbidity) and a number of different pollutants, including particulate matter (Laden et al. (2000)), sulphur dioxide (Schwartz (1991)), nitrogen dioxide (Zmirou et al. (1998)),

carbon monoxide (Conceicao et al. (2001)), and ozone (Verhoeff et al. (1996)). The most common of these associations is with particulate matter, which has been estimated for a number of metrics including PM_{10} (Samet et al. (2000)), $PM_{2.5}$ (Goldberg et al. (2001)), TSP (Lee et al. (2000)), Black Smoke (Verhoeff et al. (1996)), and CoH (Gwynn et al. (2000)). One aim of current research is to identify the component of this pollutant that is harmful to human health, and while Gwynn et al. (2000) argue for aerosol acidity (H^+), Buckeridge et al. (2002) identify the role of fine particles ($PM_{2.5}$). However measures of individual pollutants are likely to be highly correlated (for example Lin et al. (1999) find a correlation of 0.73 between PM_{10} and SO_2), meaning that the association of interest may be a by-product of one with a different pollutant. A naive solution is to include multiple pollutants in a single model, but this may cause collinearity (using linear models) or concurvity (using additive models, see Ramsay et al. (2003)) problems reducing the accuracy of the estimated associations. These problems can be overcome by undertaking a sensitivity analysis, that applies a number of single and multi pollutant models to each data set. For example Samet et al. (2000) investigate the associations between total non-accidental mortality, ozone and PM_{10}, and find that the estimated association with PM_{10} is invariant to including ozone in the model.

In contrast interest may lie in estimating the association between mortality and overall 'pollution exposure', as opposed to with specific pollutants. Yu et al. (2000) estimate a joint association for particulate matter and carbon monoxide, by calculating a relative risk for an increase in both pollutants. They report that although the individual relative risks are much smaller, the joint risk is similar in magnitude to those found in single pollutant analyses. In contrast Hong et al. (1999) define overall pollution indices, which are the sum of individual pollutants scaled by their mean. They report larger relative risks using these overall indices than were found from corresponding single pollutant models.

3.3.2 Lag

The health problems that result from pollution exposure may be felt immediately, that is on the same day (see for example Moolgavkar (2000)), after a lag of one or two days (see for example Peters et al. (2000)), or from continued exposure over the preceding few days (see for example Lin et al. (1999)). The model in equation (3.1) estimates the first of these, whilst the latter two can be estimated by replacing w_i with w_{i-1} and $\tilde{w}_i = (1/r)\sum_{j=0}^{r-1} w_{i-j}$ respectively, where r represents the length of the averaging period. The choice between different lags and averaging periods is a longstanding research problem, and there is little consensus over which should be used. It has been determined by numerous approaches, including selecting the lag that is the most significant, the one used by a previous study, or the one that minimises an objective criteria (such as DIC). Alternatively numerous researchers (see for example Burnett et al. (1994)) report results at multiple lags, using

both single and multi lag models. However neither of these are an adequate solution, because consecutive lags of a single pollutant are likely to be highly correlated. Therefore the association from a single lag model may be the byproduct of a true association with a different lag, while multi-lag models may be susceptible to collinearity problems reducing the accuracy of their estimates. One solution is to use distributed lag models (Almon (1965)), which have been used in air pollution studies by Schwartz (2000) and Zanobetti et al. (2000). These models include all lags from zero up to a specified maximum, and remove the effects of collinearity by constraining the associated regression parameters to follow a parametric function of lag number. For example Zanobetti et al. (2000) constrain the parameters to follow a regression spline of order k, in which the linear predictor is given by

$$\begin{aligned} \ln(\mu_i) &= \sum_{l=0}^{q} w_{i-l}\gamma_l + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\alpha}, \\ \gamma_l &= \sum_{j=0}^{k} \psi_j l^j + \sum_{r=1}^{c} \nu_r (l-\phi_r)_+^k. \end{aligned} \tag{3.3}$$

In the equation above m_+ is equal to m if $m > 0$, and is zero otherwise. Schwartz (2000) proposes a similar approach which replaces the regression spline constraint with a polynomial of order k. Both authors report that air pollution has a detrimental effect on health for a number of days after exposure, with the greatest effect being associated with exposure on the same day.

3.3.3 Pollution-health relationship

The shape of the relationship between air pollution and health has been the subject of much research, including recent reviews by Dominici et al. (2003) and Pope and Dockery (2006). Early studies (see for example Schwartz (1991) and Spix et al. (1993)) typically specified a constant relationship, which was represented by a single parameter such as γ in (3.2). However as generalised additive models became increasingly popular researchers relaxed this restriction, and allowed the association to depend on the underlying pollution level, termed a 'dose-response' relationship. This replaces $w_i\gamma$ by an arbitrary function $f(w_i|\lambda)$, whose shape and smoothness are estimated from the data. This function is most commonly modelled with regression splines, smoothing splines or LOESS smoothers, and an early use of this methodology by Schwartz (1994b) found the association increased as the pollution level rose.

In recent studies both constant and dose-response relationships have been estimated, the choice of which depends on the aims of the study. Constant associations are often estimated in epidemiological studies (see for example Mar et al. (2000) and Lin et al. (2002)),

where the overall size of the relationship is of primary interest. In contrast dose-response curves allow researchers to investigate whether the relationship between air pollution and health has a threshold level, below which no adverse effects are felt. Determining whether such a level exists is important for regulatory purposes, because air pollution legislation such as the Clean Air Act (1990) and the UK Air Quality Strategy (2000) set 'safe' levels for a number of common pollutants. The existence of a threshold level has been extensively investigated by Schwartz and Zanobetti (2000), Dominici et al. (2002) and Daniels et al. (2004), who estimate city specific, regional and national dose response relationships, using meta-smoothing (Schwartz and Zanobetti (2000)) and Bayesian hierarchical models (Dominici et al. (2002) and Daniels et al. (2004)). They analyse data from ten, twenty and a hundred U.S. cities, and report that although the relationships exhibit substantial heterogeneity between cities, the national dose-response curve is linearly increasing (with increasing air pollution) exhibiting no threshold level. However some of the city specific estimates are not monotonically increasing, suggesting that at some levels increasing pollution exposure decreased its harmful effect. These unsatisfactory results led Roberts (2004) to propose 'biologically plausible' dose-response functions, which constrain the curve to be monotonically increasing. He replaced $f(w_i|\lambda)$ by

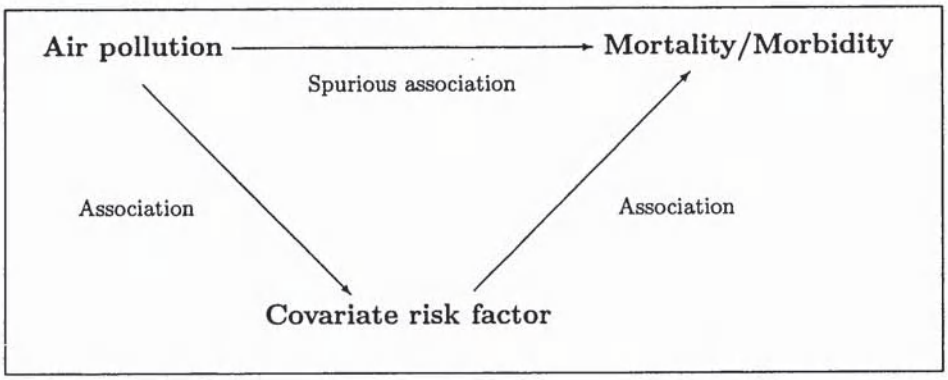
$$f(w_i|\boldsymbol{\theta},\xi) = \begin{cases} \theta_1 w_i & \text{if } w_i < \xi \\ \theta_1 \xi + \theta_2(w_i - \xi) & \text{if } w_i \geq \xi \end{cases}, \tag{3.4}$$

in which $\boldsymbol{\theta} = (\theta_1, \theta_2)$ are constrained to be positive and ξ is called a change point. However this representation is overly restrictive, because the dose-response curve only has two slopes and the point of change is fixed prior to estimation. An earlier alternative was proposed by Cakmak et al. (1999), who replaced equation (3.4) with a linear threshold model, $\theta_1(w_i - \xi)_+$, where ξ is estimated from the data. However this model is similarly restrictive, and further research is required to produce flexible alternatives that are biologically plausible. In addition to constant and dose-response relationships, a small number of studies have investigated whether the association exhibits any temporal variation. This is one of the central themes of this thesis, and is discussed in chapters four and five within the context of dynamic generalised linear models and time-varying coefficient models.

3.4 Covariate specification

The daily counts of mortality or morbidity depend on a series of covariate risk factors other than air pollution, whose influence should be removed by the covariate component of the model. If this does not happen then the estimated pollution-mortality association may be biased, an example of which is given in Figure 3-1. The covariate risk factor shown in this figure has a positive association with both air pollution and mortality, and if omitted from the model the pollution-mortality association will be positive even if no

Figure 3-1: The effect of omitting a covariate risk factor (taken from Wald (2004)).



such association exists. This phenomenon is known as confounding, and can be partially checked by examining the standardised residuals $r_i = (y_i - \mu_i)/\sqrt{\text{Var}[y_i]}$. If they resemble random fluctuations with a fixed variance confounding is less likely, because an omitted risk factor may induce structure into this series. Such structure may take the form of a long-term trend or temporal correlation, and can be diagnosed with graphical plots of the residuals. The cause of this structure maybe unknown or due to an omitted covariate such as temperature, and the methods used to remove structure from both sources is discussed below.

3.4.1 Modelling known risk factors

Measured risk factors are typically categorical or meteorological, and a wide variety of both have been used in air pollution studies. Categorical risk factors are typically represented by indicator variables, and have included 'day of the week' (see for example Kelsall and Zeger (1999)), influenza epidemics (see for example Peters et al. (2000)) and public holidays (see for example Schwartz et al. (2001)), while meteorological covariates encompass temperature (Mar et al. (2000)), dewpoint temperature (Roberts (2004)), humidity (Lee et al. (2000)), precipitation (Spix et al. (1993)), and pressure (Vedal et al. (2003)). The most common of these is temperature, because it models part of the seasonal variation typically present in health data of this type. Although meteorological covariates are routinely available their inclusion in a model requires two additional choices: (i) which lag should be used; (ii) what shape should its relationship with health take. There is little consensus over which lag of a meteorological covariate should be used, with temperature being included at lags of zero (Moolgavkar et al. (1995)), up to two days (Verhoeff et al. (1996)) or as a moving average over a number of consecutive lags (Dominici et al. (2000)). A number of researchers solve this problem by adopting the approach of Lumley and Sheppard (2000), by including the lag that has the biggest association with the daily health series. Researchers have also specified a variety of shapes for the covariate-response

relationship, including constant (Conceicao et al. (2001)), a dose-response curve (Roberts (2004)), a time-varying effect (Chiogna and Gaetan (2002)) and an indicator function for extreme values (Peters et al. (2000)). In recent studies meteorological covariates have been incorporated by modelling their effects as a dose-response relationship, which has been advocated by Dominici et al. (2000) and Daniels et al. (2004). This curve is typically modelled by a smooth function with a low degrees of freedom, because a U-shaped relationship between meteorology and mortality is often observed.

3.4.2 Modelling unknown risk factors

By their very nature, unknown risk factors cannot be added to the regression model in the same way as known covariates, and removing their influence from $\mathbf{y}$ is less straightforward. Such factors can induce long-term trends, seasonal variation, over-dispersion and temporal correlation into the mortality or morbidity series, and an early review of possible solutions was presented by Schwartz et al. (1996). Early studies (Schwartz (1993) and Spix et al. (1993)) model seasonal variation with pairs of sine and cosine terms at different frequencies, which typically range between one month and two years. The same studies remove long-term trends with parametric functions of calendar time, such as quadratic or cubic polynomials. Other early approaches model these factors with indicator variables (Verhoeff et al. (1996)) and meteorological covariates (Moolgavkar et al. (1995)), which like the parametric functions described above are overly restrictive. This restrictiveness is due to their rigid parametric nature, which lack the necessary flexibility to model excessive variation in $\mathbf{y}$. For example the sinusoidal terms force the peak in mortality to occur at the same time each year, while the monthly indicator variables do not allow for within month variation.

More recent air pollution studies model unmeasured risk factors with a smooth function of calendar time, which can be more flexible than the fixed parametric alternatives described above. Such functions have been implemented using parametric and non-parametric methods, including regression splines (Daniels et al. (2004)), smoothing splines (Dominici et al. (2000)), and locally weighted smoothers such as LOESS (Schwartz et al. (2001)). Both parametric and non-parametric formulations have advantages, with the former being easier to incorporate into a Bayesian analysis while the latter can model larger amounts of non-linear variation. General comparisons of smooth functions were carried out by Ramsay et al. (2003) and Peng et al. (2006), who compare parametric and non-parametric representations via a simulation study. In their work each representation is applied to health data with differing amounts of non-linear variation, and they report that all the methods performed well with none being preferable.

Numerous alternative methods for modelling unmeasured confounding have also been pro-

posed, including the Shumway filter which removes the long-term trend by multiplying the daily Poisson mean by

$$\zeta_i = \sum_{j=-9}^{9} \psi_j \frac{y_{i-j}}{\bar{y}}.$$

Here $\bar{y}$ denotes the mean of the health data, while the weights, $\{\psi_0, \ldots, \psi_9\}$, are symmetric and sum to one. The filter removes any slow moving temporal trends in the mortality series, including seasonal and yearly effects. Kelsall and Zeger (1999) propose an approach based in the frequency domain using Fourier transforms, while Chiogna and Gaetan (2002) use an autoregressive process within a dynamic generalised linear model. The latter approach is considered further in chapter four, where the efficacy of analysing these data using Bayesian dynamic generalised linear models is investigated. The remainder of this chapter discusses the limitations of the standard approach to modelling these data, in terms of the standard model, the data that are available, and the conclusions that can be drawn from such analyses. The next two sections describe the related problems of over-dispersion and temporal correlation.

3.5 Modelling over-dispersion

If all risk factors that influence the daily health series are known and included in the regression model, the residual variation would be adequately described by the Poisson assumption. However a subset of these risk factors are unknown, and their exclusion from the model inflates the residual variance which may make the Poisson assumption untenable. This causes the confidence intervals to be too narrow, which may falsely suggest that the pollution-health association is statistically significant. This phenomenon of increased variation in the data compared with that specified by the assumed probability distribution is known as over-dispersion, whereas under-dispersion occurs if the variation is less. The presence of over-dispersion can be checked by examining the standardised residuals r_i, which have unit variance if over-dispersion is not present. In air pollution studies over-dispersion is typically removed by adequately modelling the external risk factors, using the methods discussed in the previous section. However for some mortality series this is not sufficient, and methods that explicitly model over-dispersion are required. These approaches do not alter the estimates of the regression parameters, but inflate the corresponding standard errors to allow for this additional variation. Two standard methods are based on the quasi-likelihood and latent processes, both of which are outlined below.

3.5.1 Quasi-likelihood methods

When researchers model over-dispersion they generally use quasi-likelihood methods, which can be applied within linear (see for example Mar et al. (2000) and Peters et al. (2000)) and additive (see for example Schwartz (1994b) and Moolgavkar (2000)) regression models. The quasi-likelihood approach does not assign a parametric distribution to the response data, but only specifies its mean-variance relationship. Inference is based on the quasi-likelihood

$$\mathrm{Q}(\boldsymbol{\beta}|\mathbf{y}) = \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - k}{\phi \mathrm{V}(k)} dk,$$

in which V is the variance function of the true likelihood. The excess variation is modelled by the dispersion parameter ϕ, and for Poisson data the variance of y_i is relaxed to $\phi\mu_i$. The quasi-likelihood is derived by approximating the score function for y_i with $\frac{y_i - \mu_i}{\phi \mathrm{V}(\mu_i)}$, which have identical means, variances and expected derivatives. Incorporating over-dispersion scales the variance matrix of $\boldsymbol{\beta}$ by the estimated dispersion parameter

$$\hat{\phi} = \frac{1}{n-q} \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mathrm{V}(\mu_i)},$$

which widens the corresponding confidence intervals. Although a standard technique in likelihood based inference, quasi-likelihood methods cannot be applied in a Bayesian setting because a proper likelihood is required. A comprehensive description of quasi-likelihood is given by McCullagh and Nelder (1989).

3.5.2 Latent process methods

An alternative to quasi-likelihood are latent process methods, which model over-dispersion through the addition of a stochastic process to equation (3.2). They can be applied within both Bayesian and likelihood settings, and a simple example is the generalised linear mixed model (Breslow and Clayton (1993))

$$\begin{aligned} y_i &\sim \text{Poisson}(\epsilon_i \mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= w_i \gamma + \mathbf{z}_i^{\mathrm{T}} \boldsymbol{\alpha}, \\ \epsilon_i &\sim \mathrm{N}(1, \sigma_\epsilon^2). \end{aligned} \tag{3.5}$$

The addition of $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ relaxes the marginal mean-variance relationship to $\mathbb{E}[y_i] = \mu_i$ and $\mathrm{Var}[y_i] = \mu_i + \sigma_\epsilon^2 \mu_i^2$, which allows for over-dispersion but not under-dispersion. A similar model was used by Gwynn et al. (2000), who assume $\boldsymbol{\epsilon}$ are Gamma random variables which causes the marginal distribution of y_i to be negative binomial (Booth et al. (2003)). Other alternatives for modelling over-dispersion include assuming y_i comes from

generalised Poisson (Consul and Jain (1973)), Gaussian (Schwartz and Marcus (1990)) or Poisson double exponential family (Efron (1986)) distributions.

3.6 Modelling temporal correlation

The daily counts of mortality or morbidity form a time series, and values on successive days are likely to be correlated. This correlation is induced by external risk factors that have similar values for days close in time. If all such factors were known and included in the regression model the standardised residuals would be independent, meaning that model (3.2) would be adequate. However as previously mentioned a subset of these risk factors are unknown, and their exclusion from the model induces temporal correlation into the residuals which should be modelled. The presence of correlation can be checked by examining the autocorrelation function of the standardised residuals, with excessively large values indicating correlation is present. The majority of researchers attempt to remove this correlation by adequately modelling the external risk factors, using the methods described in section 3.4. However for some mortality series this is not sufficient, and methods that explicitly remove temporal correlation are required. Two standard approaches are based on parameter and observation driven models, a brief outline of which is given here whilst comprehensive treatments are presented by Diggle et al. (1994) and Davis et al. (1999).

3.6.1 Parameter driven methods

Parameter driven methods model temporal correlation by adding a correlated latent process to equation (3.2), and were introduced by West et al. (1985) and Zeger (1988). In air pollution studies three main types of parameter driven model have been used, which differ in how the process is incorporated into the standard model. Conceicao et al. (2001) adopt the approach of Zeger (1988), who uses a model similar to that in equation (3.5) with the random effects replaced with an autoregressive process. The mean and variance of this model are identical to those of (3.5) while the correlation is given by

$$\text{Corr}[y_i, y_{i+h}] = \frac{\rho_\epsilon(h)}{\sqrt{[1 + (\sigma_\epsilon^2 \mu_i)^{-1}][1 + (\sigma_\epsilon^2 \mu_{i+h})^{-1}]}},$$

where $\rho_\epsilon(h)$ is the correlation function of the autoregressive process. However $\text{Corr}[y_i, y_{i+h}] \leq \text{Corr}[\epsilon_i, \epsilon_{i+h}]$, making the correlation in the latent process difficult to estimate as it may be masked by little or no correlation in $\mathbf{y}$ (for further details see Davis et al. (2000)). Schwartz et al. (1996) and Vedal et al. (2003) adopt a similar approach to that of Conceicao et al. (2001), but assume an additive rather than multiplicative relationship between μ_i and ϵ_i. They model the health series with

$$
\begin{aligned}
y_i &= \mu_i + \epsilon_i \quad \text{for } i = 1, \ldots, n, \\
\ln(\mu_i) &= w_i\gamma + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\alpha}, \\
\epsilon_i &\sim \mathrm{N}(\theta_1\epsilon_{i-1} + \ldots, \theta_p\epsilon_{i-p}, \sigma_\epsilon^2),
\end{aligned}
$$

a generalised linear model with time series errors. Both models can be implemented in a Bayesian setting via MCMC simulation and in a likelihood framework using generalised estimating equations (Liang and Zeger (1986)), and further details are given elsewhere. A further modification was proposed by Chiogna and Gaetan (2002), who model these data with a dynamic generalised linear model (see chapter 2.6). This is different from the other approaches because the latent process is on the linear predictor scale, rather than being additively or multiplicatively related to y_i. Alternative parameter driven approaches for modelling count data have been proposed by Wang and Puterman (1999) and Shaddick et al. (2007), which are based on Markov chains and a sums of latent variables respectively.

3.6.2 Observation driven methods

Observation driven methods were introduced by Zeger and Qaqish (1988), and model temporal correlation by including past functions of the response variables as additional covariates. These models have been widely investigated and are also known as generalised autoregressive (Fahrmeir and Tutz (2001)) or Markov (Diggle et al. (1994)). A general form is given by

$$
\begin{aligned}
y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\
\ln(\mu_i) &= w_i\gamma + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\alpha} + \sum_{j=1}^{p} g_j(D_i)\psi_j,
\end{aligned}
$$

in which ψ_j are unknown parameters, g_j are known functions, and D_i denotes the set of past response variables and covariates. Zeger and Qaqish (1988) propose setting $g_j(D_i) = \ln(y_{i-j}^*) - w_{i-j}\gamma - \mathbf{z}_{i-j}^{\mathrm{T}}\boldsymbol{\alpha}$, the jth past residual, where $y_i^* = \max(y_i, c)$ for $c \in (0, 1)$ which prevents the occurrence of $\ln(0)$. This model has been used in an air pollution context by Xu et al. (1995), has an approximately correct marginal mean ($\mathbb{E}[y_i] \approx \mu_i$), and allows for positive and negative autocorrelation. However the resulting process is not stationary, and Davis et al. (2003) propose an extension that replaces $\{g_j\}_{j=1}^p$ with an infinite sum of scaled residuals which may result in a stationary process. Estimation for observation driven models can be implemented in both Bayesian and likelihood settings, although the references above adopt likelihood based inference using generalised estimating equations. In addition, Brumback et al. (2000) unify parameter and observations driven models into a single framework called transitional regression models, and propose a generalised least

squares approach to estimation.

3.7 Combining results from multiple locations

A range of pollution-mortality relationships have been reported in cities throughout the world, which have varied in terms of the pollutant analysed and the strength of the estimated association. A proportion of this variation is induced by researchers, who have analysed different types of data using a variety of statistical models. However the majority is caused by factors beyond human control, including the mixture of pollutants in the atmosphere, climatic conditions, structure of the local population, and availability of appropriate data. Consequently studies presenting single city analyses provide very limited evidence of the relationship between air pollution exposure and mortality. Slightly more conclusive evidence can be obtained by pooling data from multiple studies using meta analytic techniques, and early examples are presented by Schwartz (1994a), and Dockery and Pope (1994). These authors combine the results from numerous studies via a two stage random effects model, which estimates an overall pollution-mortality association as a weighted average of the city specific estimates. For example Dockery and Pope (1994) combine studies that investigate the role of particulate matter, and report consistent associations with a number of mortality and morbidity classifications. However meta-analyses are based on single city studies conducted in variable conditions, meaning that the results are affected by the quality of the individual studies.

An alternative approach is to conduct multi-city studies, which estimate associations at multiple locations using a standard protocol. This protocol is a detailed description of the data and statistical model to be used, meaning that the results from multiple cities are more comparable. An early example is Air Pollution and Health: A European Approach (APHEA, Schwartz et al. (1996), Katsouyanni et al. (1996) and Zmirou et al. (1998)), which estimates associations for multiple pollutants across ten European cities. This was followed by an extension to 29 European cities called APHEA-2 (Katsouyanni et al. (2001) and Zanobetti et al. (2002)), and a 90 city study in the USA called the National Morbidity, Mortality and Air Pollution Study (NMMAPS, Samet et al. (2000) and Daniels et al. (2004)). These studies typically exhibit less heterogeneity than a collection of single city studies, which is probably due to the consistent approach to analysis. For example the NMMAPS study reports consistent associations between particulate matter and mortality, with relative risks in the range 0.5% to 1.5%. However associations with gaseous pollutants are much less consistent, which is probably caused by factors beyond human control. Such factors are unavoidable, and multi-city studies are the best framework for obtaining comparable results from different locations. They are also an ideal basis for meta-analyses (see for example Dominici et al. (2000) and Samet et al. (2000)), which can be implemented within a Bayesian hierarchical framework using MCMC simulation. However this approach

can be computationally impractical, and some researchers (see for example Dominici et al. (2000) and Dominici et al. (2002)) conduct the single city and meta-analytic analyses separately.

3.8 Individual and ecological regression models

As described in section 3.1, the health data are usually only available as aggregated daily counts, meaning that an ecological regression model is required. Such models answer fundamentally different epidemiological questions from individual level models (for a review see Plummer and Clayton (1996)), and the results should not be stated in terms of a causal link between air pollution and health. The use of ecological studies in this context is contentious, and has been discussed by Richardon et al. (1987), Greenland and Morgenstern (1989) and Wakefield and Salway (2001), the latter of which describes their use as '*appealing, at least for hypothesis generation*'. A causal relationship between air pollution and health can only be estimated from individual level data, but personal mortality or morbidity events cannot be obtained for confidentiality reasons, while pollution exposures are expensive and impractical to obtain for more than a few individuals. A few researchers (see for example Lioy et al. (1990) and Ozkaynak et al. (1996)) have collected individual pollution exposure data, but it has been limited by cost to less than ten individuals per day for 50 consecutive days.

As such data are largely unavailable, a small number of researchers (see for example Zeger et al. (2000), Sheppard and Damian (2000), Wakefield and Shaddick (2006)) have attempted to estimate a causal link between air pollution and health by aggregating individual exposure-response models to the population level. However such individual level models are typically nonlinear (based on Bernoulli observations), meaning that the aggregation cannot be done exactly. The resulting error is known as ecological bias, and a review is given by Wakefield and Salway (2001). Zeger et al. (2000) adopt an individual level Bernoulli risk model, and aggregate it to the population level using a Taylor series approximation. They relate unknown personal exposures to ambient concentrations via a measurement error approach, which includes both classical and Berkson sources of error. However their approximation depends on an average of personal exposures, limiting its use in practice. A similar approach was presented by Sheppard and Damian (2000), who incorporate an individual exposure distribution based on ambient and indoor pollution sources. However, in common with the approach of Zeger et al. (2000) the data required to fit the full model is unavailable, and simplifying assumptions are made to compensate. In contrast Wakefield and Shaddick (2006) attempt to avoid ecological bias by averaging the individual level risks, rather than estimating the risk from the average exposure.

3.9 Measuring air pollution exposure

As the health data are ecological, the desired exposure is the average level of pollution experienced by the population, which is realistically unobtainable in practice. Consequently researchers estimate pollution by averaging the measurements over the k monitoring sites, which is denoted by $\mathbf{w}$ in section 3.2. These data represent ambient pollution levels rather than personal or population exposures, and are likely to be a poor estimate of the latter in three main ways. Firstly air pollution originates from indoor (for example from gas cookers) as well as outdoor sources, and as the population spend a significant proportion of their time indoors, pollution exposure may be poorly estimated by ambient levels. Secondly the ambient monitors are likely to measure with error, meaning that the daily averages may be inaccurate. Thirdly the underlying pollution field is likely to exhibit spatial variation, meaning that measurements at a small number of sites are likely to be a poor approximation of average population exposure. This is because the population live and work throughout the region under study, and do not all spend their lives close to a monitoring site. This is exacerbated by the locations of these monitors, which are unlikely to be equally spaced throughout the region under study. Instead they are strategically positioned, for example at pollution hot spots which allows the highest levels to be monitored. As a result researchers have proposed alternative methods for estimating air pollution exposure, which have followed two separate approaches. The first attempts to estimate population exposure rather than ambient levels, while the second aims to obtain more realistic estimates of ambient levels.

Dominici and Zeger (2000) estimate average population exposure using a measurement error approach, that is based on small amounts of personal exposure data (see for example Lioy et al. (1990) and Ozkaynak et al. (1996)). They adopt a two stage modelling strategy, that relates the health data to the unknown population exposure at the first stage. The second stage is a measurement error model, which relates ambient concentrations to average population exposure in five external data sets. These additional data include ambient levels and personal exposures from other populations, and Dominici and Zeger (2000) transport the relationship between these data to the city under study. In contrast Burke et al. (2001) and Zidek et al. (2005) generate large samples of personal exposures using an exposure simulator. Their simulators are known as SHEDS-PM and pCNEM respectively, and generate daily personal exposures based on ambient levels and time activity patterns for members of the population. Holloman et al. (2004) use the exposures generated by SHEDS-PM as an estimate of average population exposure, which they relate to daily mortality via a Bayesian measurement error model. In common with Dominici and Zeger (2000) they report increased relative risks compared with using ambient pollution data.

However the relationship between ambient pollution levels and mortality is of interest in

its own right, because current legislation such as the Clean Air Act (1990) and the UK Air Quality Strategy (2000) are based on ambient data. As described in section 3.2 spatially averaged ambient pollution levels are estimated by $\mathbf{w}$, which is likely to be a poor estimate because it does not account for measurement error or spatial variation in the observed data. In chapter six I investigate how measurement error and spatial variation affect the validity of $\mathbf{w}$, and propose two alternative estimates that incorporate these factors.

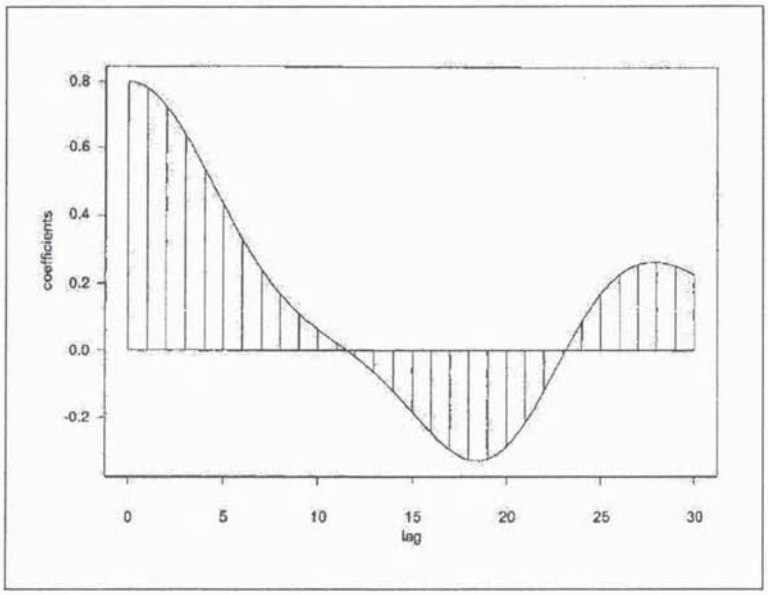
3.10 Mortality displacement

The significance of air pollution in terms of public health depends on which part of the population suffer ill health as a result of exposure. One commonly espoused viewpoint is that air pollution exposure only affects frail individuals, whose remaining life expectancy would have been short regardless of their pollution exposure. In other words air pollution exposure only brings death forward by a few days for frail individuals, whose quality of life in these final days would have been limited. This is known as '*mortality displacement*' or the '*harvesting hypothesis*', and if it is true the public health significance of air pollution would be dramatically reduced. This hypothesis has been the subject of recent research, and a brief review is given here.

Zeger et al. (1999) and Schwartz (2001) investigate mortality displacement by estimating the pollution-mortality association at different time scales. If such associations only occur at the shortest time scales, then pollution only advances death by a few days and mortality displacement is supported. Conversely, associations at longer time scales indicate that pollution exposure is likely to cause a significant loss of life, which increases its public health significance. Both authors propose pollution-mortality estimates that are resistant to mortality displacement, using frequency and time domain approaches. Zeger et al. (1999) adopt the frequency domain approach of Kelsall and Zeger (1999), which replaces the data with their discrete fourier transforms. They estimate the overall relationship between air pollution and mortality as a weighted average of frequency specific associations, giving high frequencies (likely to be caused by mortality displacement) zero weight. In contrast Schwartz (2001) adopts a time domain approach, which is based on applying linear filters to the data. He removes any short-term variation by calculating moving averages between 15 and 60 days, which are then analysed using a standard Poisson model. Both authors report positive associations between air pollution and health, suggesting that the mortality displacement hypothesis is not supported.

Zanobetti et al. (2000) adopt an alternative approach to investigating mortality displacement based on distributed lag models (see equation (3.3)). Such models estimate the association between air pollution and health at a number of lags, and they claim that mortality displacement manifests itself in a pattern similar to Figure 3-2 (taken from Zanobetti et al.

Figure 3-2: The likely lag structure under the mortality displacement hypothesis (taken from Zanobetti et al. (2002)).



(2002)). They argue that a day of high air pollution will cause a large number of deaths in the pool of frail individuals, depleting its size and leading to lower mortality on subsequent days (which induces the negative associations seen in Figure 3-2). Zanobetti et al. (2002) investigate the possibility of mortality displacement by applying distributed lag models to data from ten European cities. These data are part of the APHEA-2 study, and the results are inconsistent with mortality displacement. However Murray and Nelson (2000) investigate this hypothesis using a modified dynamic generalised linear model, and report results that are consistent with mortality displacement. The evidence for and against this phenomenon is summarised by Dominici et al. (2003), who argue that the differences between the results may be caused by long-term confounding and as such requires further investigation.

Chapter 4

Modelling air pollution and health data using Bayesian dynamic generalised linear models

4.1 Introduction

In this chapter I model air pollution and health data using a Bayesian dynamic generalised linear model (DGLM, West et al. (1985)), and compare the efficacy of this approach against more traditional alternatives. Dynamic generalised linear models were introduced in chapter 2.6, and extend generalised linear models by allowing the regression parameters to evolve over time via an autoregressive process. This process can take on a number of different forms, including a set of independent univariate processes for individual regression parameters, a multivariate correlated process for a block of parameters, or parameters that are constant over time. The only known analysis of air pollution and health data using a DGLM is that of Chiogna and Gaetan (2002), and this work differs from theirs in that a Bayesian approach is adopted with estimation based on Markov chain monte carlo simulation. In addition I provide a more in-depth comparison with the standard approaches to modelling these data outlined in chapter three.

The remainder of this chapter is organised as follows. Section 4.2 reviews the background and motivation for this work, and describes the impact that it may have for future air pollution and health studies. In particular the efficacy of dynamic models in this context is discussed, and two extensions to the standard model are proposed. Section 4.3 describes the Bayesian dynamic generalised linear model proposed in this chapter, and provides a brief comparison with the model used by Chiogna and Gaetan (2002). Section 4.4 describes the Markov chain monte carlo algorithm adopted in this chapter, and reviews alternative schemes that have previously been proposed. Section 4.5 presents a case

study investigating the efficacy of dynamic models in this context, by analysing data from Greater London. Finally section 4.6 presents a concluding discussion.

4.2 Background and motivation

Dynamic generalised linear models have had limited use in air pollution and health studies, and the only known analysis is that of Chiogna and Gaetan (2002). In that study a dynamic model is used to estimate the relationship between total non-accidental mortality and exposure to PM_{10} in Birmingham Alabama, and their model has the general form

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_i, \\ \boldsymbol{\beta}_i &\sim \mathrm{N}(\Phi \boldsymbol{\beta}_{i-1}, \Sigma_\beta), \\ \boldsymbol{\beta}_0 &\sim \mathrm{N}_q(\boldsymbol{\mu}_0, \Sigma_0). \end{aligned} \tag{4.1}$$

The $q \times 1$ vector of explanatory variables $(\mathbf{x}_i)$ includes a measure of air pollution and a vector of covariates, and their model contains three independent autoregressive processes. The underlying trend in the mortality data is modelled by a second order random walk, while the effects of PM_{10} and humidity are allowed to evolve over time as first order random walks. Chiogna and Gaetan adopt the likelihood approach to estimation proposed by Fahrmeir and co-workers (see Fahrmeir and Kaufmann (1991), Fahrmeir (1992) and Fahrmeir and Wagenpfeil (1997)), which is based on an iterative algorithm involving the Kalman filter and generalised cross validation. Further details of the algorithm are given in chapter 2.6, whilst a full description is given by Fahrmeir and Wagenpfeil (1997). In this chapter I model air pollution and health data using a Bayesian dynamic generalised linear model, and the analysis is motivated by two aims. Firstly the Bayesian paradigm is the natural setting in which to view hierarchical models of this type, and I compare the Bayesian approach adopted here with the likelihood based alternative used by Chiogna and Gaetan (2002). Secondly dynamic models allow the standard approaches to modelling these data to be extended in two ways, both of which should be investigated. These aims are discussed in more detail in sections 4.2.1 and 4.2.2, and are investigated in the case study in section 4.5.

4.2.1 Comparison between Bayesian and likelihood estimation

The Bayesian approach proposed here has a number of advantages over the likelihood based analysis presented by Chiogna and Gaetan. In a Bayesian approach the posterior distribution of $\boldsymbol{\beta}$ correctly allows for the variability in (Σ_β, Φ), while confidence intervals calculated within a likelihood setting do not. In a likelihood analysis (Σ_β, Φ) are estimated

by optimising a data driven criterion, meaning that the standard errors of β are calculated assuming (Σ_β, Φ) are fixed at their estimated values. As a result the confidence intervals for β are likely to be too narrow, which may lead to a statistically insignificant effect of air pollution appearing to be significant. In contrast the Bayesian credible intervals are likely to be the correct width, because $(\beta, \Sigma_\beta, \Phi)$ are simultaneously estimated within the MCMC algorithm.

The Bayesian approach also allows prior knowledge of the parameters to be incorporated into the model, whilst results similar to a likelihood based analysis may be obtainable by specifying prior ignorance. This is particularly important for dynamic models, where the regression parameters are likely to evolve smoothly over time. This smoothness is controlled by (Σ_β, Φ), and non-informative priors may result in the estimated parameter process being contaminated with unwanted noise. Such noise is likely to hide any trend in the parameter process, and can be removed by specifying informative priors for (Σ_β, Φ). The Bayesian paradigm is also the natural framework in which to view hierarchical models of this type, because it can incorporate variation at multiple levels whilst making use of standard estimation techniques. However Bayesian analyses are typically time consuming, and likelihood based estimation is quicker to implement. To assess the relative performance of these approaches I apply all models in section 4.5 using the Bayesian algorithm described in section 4.4, and the likelihood based alternative adopted by Chiogna and Gaetan (2002).

4.2.2 Comparison between standard and dynamic models

As discussed in chapter 3.2 air pollution and health data are typically modelled by Poisson regression methods, in which daily counts of mortality or morbidity are regressed against air pollution levels and a vector of covariates. Analysing these data with dynamic models allows the standard Poisson model to be extended in two ways: (i) the influence of unmeasured risk factors can be removed with an autoregressive process; (ii) the effects of air pollution can evolve over time. Investigating the efficacy of these two extensions is a major aim of this work, and the motivation for both is given below.

Modelling the influence of unmeasured risk factors

Meteorological covariates generally exhibit periodic behaviour, and their inclusion in a model removes some of the seasonal variation from the health data. The remainder of this variation is typically removed by a smooth function of calendar time, which also models any long-term trends, over-dispersion and temporal correlation present in the daily health series. Analysing air pollution and health data with dynamic models allows these factors to be modelled with an autoregressive process, rather than this smooth function of calendar time. Such a process sits in discrete time and estimates the temporal trend in $\mathbf{y}$, while its smoothness is mainly controlled by a single parameter (the evolution variance). In these

respects it is a discrete time analogue of a smooth function of calendar time, and has the same discrete time support as the mortality series. In the dynamic modelling literature (see for example Chatfield (1996) and Fahrmeir and Tutz (2001)) long-term trends and temporal correlation are commonly modelled by one of three processes:

$$\begin{aligned}
\text{First order random walk} \quad \beta_i &\sim \mathrm{N}(\beta_{i-1}, \tau^2), \\
\text{Second order random walk} \quad \beta_i &\sim \mathrm{N}(2\beta_{i-1} - \beta_{i-2}, \tau^2), \\
\text{Local linear trend model} \quad \beta_i &\sim \mathrm{N}(\beta_{i-1} + \delta_{i-1}, \tau^2), \\
\delta_i &\sim \mathrm{N}(\delta_{i-1}, \psi^2).
\end{aligned} \tag{4.2}$$

All of these processes are non-stationary which allows their underlying mean level to change over time, a desirable characteristic when modelling long-term trends. A second order random walk is the natural choice from these alternatives, because it is the discrete time analogue of a natural cubic spline of calendar time (Fahrmeir and Tutz (2001)), one of the standard approaches for representing the smooth functions. Chiogna and Gaetan (2002) use a second order random walk for this reason, and in section 4.5 I extend their work by comparing the performance of smooth functions with each of the three processes listed above. I estimate the smooth function with a natural cubic spline because it is parametric, making estimation within a Bayesian setting straightforward.

Investigating the shape of the air pollution and health relationship

The pollution-mortality association is typically assumed to be constant or depend on the pollution level, the latter of which replaces $w_i\gamma$ in (3.2) with a smooth function $f(w_i|\lambda)$. This is called a dose-response relationship, and higher pollution levels typically result in larger adverse effects. Comparatively little research has allowed this association to exhibit temporal variation, which is likely to be seasonal or a long-term trend. Seasonal variation may be caused by an interaction with temperature or another pollutant exhibiting a seasonal pattern, while a long-term trend may result from a slow change in the composition of harmful pollutants, or from a change in the size and structure of the population at risk. If the pollution-mortality relationship dose exhibit temporal variation, then knowing its shape would allow a better understanding of how air pollution interacts with external factors such as temperature. The only known studies that investigate the time-varying relationship between air pollution and health are those of Moolgavkar et al. (1995), Chiogna and Gaetan (2002) and Peng et al. (2005), who model these associations as:

- Moolgavkar et al - $\gamma_i = \begin{cases} \theta_1 & \text{for spring days} \\ \theta_2 & \text{for summer days} \\ \theta_3 & \text{for autumn days} \\ \theta_4 & \text{for winter days} \end{cases}$;
- Chiogna et al - $\gamma_i \sim \mathrm{N}(\gamma_{i-1}, \sigma^2)$;
- Peng et al - $\gamma_i = \theta_1 + \theta_2 \sin(2\pi i/365) + \theta_3 \cos(2\pi i/365)$.

Moolgavkar et al. (1995) and Peng et al. (2005) investigate the possibility of seasonal variation in γ, by forcing the relationship to follow a rigid parametric form. Moolgavkar et al. (1995) model this association as a step function with four discrete values (one for each season), while Peng et al. (2005) use a smooth cyclical function with a period of one year. Both these approaches are too restrictive, and do not allow the temporal variation to exhibit shapes that are not seasonal. In contrast, the first order random walk adopted by Chiogna and Gaetan (2002) does not fix the form of the time-varying relationship *a-priori*, and allows its shape to be estimated from the data resulting in a more realistic model. In section 4.5 I investigate whether the association between PM_{10} and respiratory mortality in Greater London exhibits temporal variation, by modelling it as a first order random walk. I use a first order random walk for the flexibility described above, and because it allows a comparison with the work of Chiogna and Gaetan (2002).

4.3 Bayesian dynamic generalised linear models

The Bayesian dynamic generalised linear model proposed in this chapter is a re-specification of that used by Chiogna and Gaetan (2002, see equation (4.1)) that is given by

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_i + \mathbf{z}_i^{\mathrm{T}} \boldsymbol{\alpha}, \\ \boldsymbol{\beta}_i &\sim \mathrm{N}(\Phi_1 \boldsymbol{\beta}_{i-1} + \ldots + \Phi_p \boldsymbol{\beta}_{i-p}, \Sigma_\beta), \\ \boldsymbol{\beta}_{-p+1}, \ldots, \boldsymbol{\beta}_{-1}, \boldsymbol{\beta}_0 &\sim \mathrm{N}(\boldsymbol{\mu}_0, \Sigma_0), \\ \boldsymbol{\alpha} &\sim \mathrm{N}(\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \\ \Sigma_\beta &\sim \text{Inverse-Wishart}(n_\Sigma, S_\Sigma^{-1}). \end{aligned} \tag{4.3}$$

The daily health counts are denoted by $\mathbf{y} = (y_1, \ldots, y_n)$, and the covariates include an $r \times 1$ vector $\mathbf{z}_i$ with static parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)_{r \times 1}$, and a $q \times 1$ vector $\mathbf{x}_i$ with dynamic parameters $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{iq})_{q \times 1}$. The dynamic parameters are assigned an autoregressive prior of order p, which is initialised by starting parameters $(\boldsymbol{\beta}_{-p+1}, \ldots, \boldsymbol{\beta}_0)$ at times $(-p+1, \ldots, 0)$. Each initialising parameter has a Gaussian prior with mean $\boldsymbol{\mu}_{0_{q \times 1}}$ and variance $\Sigma_{0_{q \times q}}$, and are included to allow $\boldsymbol{\beta}_1$ to follow an autoregressive process

$(\boldsymbol{\beta}_1 \sim \mathrm{N}(\Phi_1\boldsymbol{\beta}_0 + \ldots + \Phi_p\boldsymbol{\beta}_{-p+1}, \Sigma_\beta)$. The autoregressive parameters can be stacked into a single vector denoted by $\boldsymbol{\beta} = (\boldsymbol{\beta}_{-p+1}, \ldots, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n)_{(n+p)q\times 1}$, whose characteristics are controlled by the $q \times q$ variance and state matrices $(\Sigma_\beta, \Phi_1, \ldots, \Phi_p)$. The variation in $\boldsymbol{\beta}$ is largely determined by Σ_β, which is assigned a conjugate inverse-Wishart prior. For univariate processes Σ_β is scalar, and the conjugate prior simplifies to an inverse-gamma distribution. The stationarity of $\boldsymbol{\beta}$ is determined by the state matrices $\Phi = \{\Phi_1, \ldots, \Phi_p\}$, which may contain unknown parameters or known constants meaning that prior specification depends on their form. For example a univariate first order autoregressive process is stationary if $|\Phi_1| < 1$, while prior specification is discussed in section 4.4. A Gaussian prior is assigned to $\boldsymbol{\alpha}$ because prior information is simple to specify in this form. The unknown parameters are $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_\beta)$ and components of Φ, whereas the hyperparameters $(\boldsymbol{\mu}_\alpha, \Sigma_\alpha, n_\Sigma, S_\Sigma, \boldsymbol{\mu}_0, \Sigma_0)$ are known.

The model proposed above is a re-formulation of that used by Chiogna and Gaetan (see equation 4.1), which fits naturally within the Bayesian framework adopted here. Apart from the inclusion of prior distributions there are two major differences between the two models, the first of which is operational while the second is notational. Firstly a vector of covariates with static parameters are explicitly included in the linear predictor, allowing the static and dynamic parameters to be updated separately in the MCMC simulation algorithm. This enables the autoregressive characteristics of $\boldsymbol{\beta}$ to be incorporated into its Metropolis-Hastings step, without forcing the same property onto the simulation of the static parameters. This would not be possible using (4.1), because covariates with static parameters are included in the AR(1) process by a particular specification of Σ_β and Φ (diagonal elements of Σ_β and Φ are zero and one respectively). The specification in (4.1) is also inefficient, because n identical copies of each static parameter are estimated. Secondly (4.1) appears to be an AR(1) process, which compares with my more general AR(p) process. In fact an AR(p) process can be written in the form of (4.1) by a particular specification of $(\boldsymbol{\beta}, \Sigma_\beta, \Phi)$, but for notational clarity (4.3) is adopted here.

4.4 MCMC estimation algorithm

This section describes an MCMC simulation algorithm for model (4.3), and reviews alternative schemes that have been proposed in the literature. A review of non-Bayesian estimation for dynamic models is given in chapter 2.6. The joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_\beta, \Phi)$ is given by

$$
\begin{aligned}
f(\boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_\beta, \Phi|\mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}) f(\boldsymbol{\alpha}) f(\boldsymbol{\beta}|\Sigma_\beta, \Phi) f(\Sigma_\beta) f(\Phi), \\
&= \prod_{i=1}^{n} \text{Poisson}(y_i|\boldsymbol{\beta}_i, \boldsymbol{\alpha}) \text{N}(\boldsymbol{\alpha}|\boldsymbol{\mu}_\alpha, \Sigma_\alpha) \prod_{i=1}^{n} \text{N}(\boldsymbol{\beta}_i|\Phi_1\boldsymbol{\beta}_{i-1} + \ldots + \Phi_p\boldsymbol{\beta}_{i-p}, \Sigma_\beta) \\
&\times \text{N}(\boldsymbol{\beta}_{-p+1}|\boldsymbol{\mu}_0, \Sigma_0) \ldots \text{N}(\boldsymbol{\beta}_0|\boldsymbol{\mu}_0, \Sigma_0) \text{Inverse-Wishart}(\Sigma_\beta|n_\Sigma, S_\Sigma^{-1}) f(\Phi),
\end{aligned}
$$

where $f(\Phi)$ depends on the form of the AR(p) process. The next subsection outlines the overall simulation algorithm, with specific details given in subsections 4.4.2 to 4.4.5.

4.4.1 Overall algorithm

The parameters are updated using a block Metropolis-Hastings algorithm, in which starting values $(\boldsymbol{\beta}^{(0)}, \boldsymbol{\alpha}^{(0)}, \Sigma_\beta^{(0)}, \Phi^{(0)})$ are generated from over-dispersed versions of the priors (for example t-distributions replace Gaussian distributions). The algorithm alternately samples from the full conditional distributions of the following blocks.

(a) **Dynamic parameters** $\boldsymbol{\beta} = (\boldsymbol{\beta}_{-p+1}, \ldots, \boldsymbol{\beta}_n)$.
Further details are given in subsection (4.4.2).

(b) **Static parameters** $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$.
Further details are given in subsection (4.4.3).

(c) **Variance matrix** Σ_β.
Further details are given in subsection (4.4.4).

(d) **AR(p) matrices,** $\Phi = (\Phi_1, \ldots, \Phi_p)$ (or components of).
Further details are given in subsection (4.4.5).

4.4.2 Sampling from $f(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}, \Sigma_\beta, \Phi)$

The full conditional of $\boldsymbol{\beta}$ (given below) is the product of n Poisson observations and a Gaussian prior, the latter of which consists of n autoregressive distributions for $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n)$ and p initialising distributions for $(\boldsymbol{\beta}_{-p+1}, \ldots, \boldsymbol{\beta}_0)$.

$$
\begin{aligned}
f(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}, \Sigma_\beta, \Phi) &\propto \prod_{i=1}^{n} \text{Poisson}(y_i|\boldsymbol{\beta}_i, \boldsymbol{\alpha}) \prod_{i=1}^{n} \text{N}(\boldsymbol{\beta}_i|\Phi_1\boldsymbol{\beta}_{i-1} + \ldots + \Phi_p\boldsymbol{\beta}_{i-p}, \Sigma_\beta) \\
&\times \text{N}(\boldsymbol{\beta}_{-p+1}|\boldsymbol{\mu}_0, \Sigma_0) \cdots \text{N}(\boldsymbol{\beta}_0|\boldsymbol{\mu}_0, \Sigma_0)
\end{aligned}
$$

The first p parameters are updated separately from $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$, because their full conditional distribution is Gaussian and does not depend on $\mathbf{y}$. In contrast the full conditional of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ is non-standard, because it is a product of Poisson and Gaussian distributions which cannot be sampled from directly. A number of sampling algorithms have

been proposed for this problem, and a brief review is given here. Fahrmeir et al. (1992) combine a rejection sampling algorithm with a Gibbs step, but report acceptance rates that are very low making the algorithm prohibitively slow. In contrast Shephard and Pitt (1997) and Gamerman (1998) suggest Metropolis-Hastings algorithms, with proposal distributions based on Fisher scoring steps and Taylor series expansions respectively. These proposal distributions are computationally expensive to calculate, so the conditional prior proposal algorithm of Knorr-Held (1999) is used here. His proposal distribution is computationally cheap to calculate compared with those of Shephard and Pitt (1997) and Gamerman (1998), while the Metropolis-Hastings acceptance rate has a simple form and is easy to calculate. The proposal distribution is based on the prior $f(\boldsymbol{\beta})$, and ignoring $\boldsymbol{\beta}_{-p+1}, \ldots, \boldsymbol{\beta}_0$ which have already been sampled, this can be re-written as a singular multivariate Gaussian distribution:

$$\begin{aligned} f(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n | \Sigma_\beta, \Phi, \boldsymbol{\beta}_{-p+1}, \ldots, \boldsymbol{\beta}_0) &= \prod_{i=1}^{n} \mathrm{N}(\boldsymbol{\beta}_i | \Phi_1 \boldsymbol{\beta}_{i-1} + \ldots + \Phi_p \boldsymbol{\beta}_{i-p}, \Sigma_\beta), \\ &\propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}} K \boldsymbol{\beta}\right). \end{aligned}$$

The variance matrix K^{-1} does not exist, but its inverse is given by

$$K = \begin{bmatrix} K_{-p+1,-p+1} & \cdots & K_{-p+1,n} \\ \vdots & & \vdots \\ K_{n,-p+1} & \cdots & K_{n,n} \end{bmatrix}, \tag{4.4}$$

where $K_{i,i}$ is a $q \times q$ block relating to $\boldsymbol{\beta}_i$. This matrix has a bandwidth of p blocks, that is all blocks $K_{i,j}$, with $|i-j| > p$ are zero, making the computation less expensive. The exact form of K depends on the order of the AR(p) process, and formulae for first and second order processes are given below (all other blocks are zero).

First order

$$\begin{aligned} K_{i,i} &= \begin{cases} \Phi_1^{\mathrm{T}} \Sigma_\beta^{-1} \Phi_1 & i = 0 \\ \Phi_1^{\mathrm{T}} \Sigma_\beta^{-1} \Phi_1 + \Sigma_\beta^{-1} & i = 1, \ldots, n-1 \\ \Sigma_\beta^{-1} & i = n \end{cases} \\ K_{i,i+1} &= -\Phi_1^{\mathrm{T}} \Sigma_\beta^{-1} \qquad \forall\, i \\ K_{i,i-1} &= -\Sigma_\beta^{-1} \Phi_1 \qquad \forall\, i \end{aligned}$$

Second order

$$
\begin{aligned}
K_{i,i} &= \begin{cases} \Phi_2^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_2 & i=-1 \\ \Phi_1^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_1+\Phi_2^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_2 & i=0 \\ \Sigma_\beta^{-1}+\Phi_1^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_1+\Phi_2^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_2 & i=1,\ldots,n-2 \\ \Sigma_\beta^{-1}+\Phi_1^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_1 & i=n-1 \\ \Sigma_\beta^{-1} & i=n \end{cases} \\
K_{i,i+1} &= \begin{cases} \Phi_2^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_1 & i=-1 \\ -\Phi_1^{\mathrm{T}}\Sigma_\beta^{-1}+\Phi_2^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_1 & i=0,\ldots,n-2 \\ -\Phi_1^{\mathrm{T}}\Sigma_\beta^{-1} & i=n-1 \end{cases} \\
K_{i,i-1} &= \begin{cases} \Phi_1^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_2 & i=0 \\ -\Sigma_\beta^{-1}\Phi_1+\Phi_1^{\mathrm{T}}\Sigma_\beta^{-1}\Phi_2 & i=1,\ldots,n-1 \\ -\Sigma_\beta^{-1}\Phi_1 & i=n \end{cases} \\
K_{i,i+2} &= -\Phi_2^{\mathrm{T}}\Sigma_\beta^{-1} \qquad \forall\, i \\
K_{i,i-2} &= -\Sigma_\beta^{-1}\Phi_2 \qquad \forall\, i
\end{aligned}
$$

The parameters are updated in blocks of size g, which is a sensible intermediate strategy between the low acceptance rates obtained from sampling high dimensional vectors, and the computational burden of sampling variables individually (Liu et al. (1994)). The block size act as a tuning parameter, which can be altered to achieve the desired acceptance rates. The proposal distribution for a block of parameters is its conditional distribution given the remaining elements of $\boldsymbol{\beta}$. For example, for $\boldsymbol{\beta}_{r,s} = (\boldsymbol{\beta}_r, \ldots, \boldsymbol{\beta}_s)_{gq\times 1}$ $(s = r+g-1)$ this is given by

$$
f(\boldsymbol{\beta}_{r,s}|\boldsymbol{\beta}_{-r,s}, \Sigma_\beta, \Phi) \sim \mathrm{N}(\boldsymbol{\mu}_{r,s}, \Sigma_{r,s}),
$$

where $\boldsymbol{\beta}_{-r,s} = (\boldsymbol{\beta}_{-p+1}\ldots, \boldsymbol{\beta}_{r-1}, \boldsymbol{\beta}_{s+1}, \ldots, \boldsymbol{\beta}_n)$. The mean and variance of this distribution are given by

$$
\begin{aligned}
\boldsymbol{\mu}_{r,s} &= \begin{cases} -\tilde{K}_{r,s}^{-1}\tilde{K}_{-p+1,r-1}\boldsymbol{\beta}_{-p+1,r-1} & \text{if s=n} \\ -\tilde{K}_{r,s}^{-1}\tilde{K}_{s+1,n}\boldsymbol{\beta}_{s+1,n} & \text{if r=-p+1} \\ -\tilde{K}_{r,s}^{-1}(\tilde{K}_{-p+1,r-1}\boldsymbol{\beta}_{-p+1,r-1}+\tilde{K}_{s+1,n}\boldsymbol{\beta}_{s+1,n}) & \text{otherwise} \end{cases}, \\
\Sigma_{r,s} &= \tilde{K}_{r,s}^{-1},
\end{aligned}
$$

which are calculated using standard properties of the multivariate Gaussian distribution. In this calculation the precision matrix is decomposed into

$$K = \begin{pmatrix} & \tilde{K}^{\mathrm{T}}_{-p+1,r-1} & \\ \tilde{K}_{-p+1,r-1} & \tilde{K}_{r,s} & \tilde{K}_{s+1,n} \\ & \tilde{K}^{\mathrm{T}}_{s+1,n} & \end{pmatrix},$$

where $\tilde{K}_{r,s}$ is the square $gq \times gq$ matrix containing blocks $K_{r,r}$ to $K_{s,s}$. The remaining two sub-matrices are rectangular, contain the same rows as $\tilde{K}_{r,s}$, and include all the remaining columns. To avoid any mixing problems at the boundaries of each block, the length of the first block is randomly generated from the set $\{q, 2q, \ldots, gq\}$. The acceptance probability of a move from $\boldsymbol{\beta}^{(j)}_{r,s}$ to $\boldsymbol{\beta}^{*}_{r,s}$ is given by

$$\min\left\{1, \frac{\prod_{i=r}^{s} \text{Poisson}(y_i|\boldsymbol{\beta}^{*}_{i}, \boldsymbol{\alpha}^{(j)})}{\prod_{i=r}^{s} \text{Poisson}(y_i|\boldsymbol{\beta}^{(j)}_{i}, \boldsymbol{\alpha}^{(j)})}\right\},$$

and further details are given by Knorr-Held (1999).

4.4.3 Sampling from $f(\boldsymbol{\alpha}|\mathbf{y}, \boldsymbol{\beta}, \Sigma_\beta, \Phi)$

The full conditional of $\boldsymbol{\alpha}$ (given below) is the product of n Poisson observations and a multivariate Gaussian prior.

$$f(\boldsymbol{\alpha}|\mathbf{y}, \boldsymbol{\beta}) \propto \prod_{i=1}^{n} \text{Poisson}(y_i|\boldsymbol{\beta}_i, \boldsymbol{\alpha}) \text{N}(\boldsymbol{\alpha}|\boldsymbol{\mu}_\alpha, \Sigma_\alpha) \tag{4.5}$$

The Gaussian prior is not conjugate to the Poisson data, which results in a non-standard full conditional distribution. A Metropolis-Hastings algorithm is typically used for updating $\boldsymbol{\alpha}$, and two common proposal distributions are based on random walks and Fisher scoring steps (Fahrmeir and Tutz (2001)). Throughout this thesis I adopt a random walk proposal for updating static regression parameters, because it is computationally cheaper to calculate and has a tuning parameter to control the acceptance rates. The parameters are updated in blocks for the reasons previously discussed, and the proposal distribution is given by $\text{N}(\boldsymbol{\alpha}^{(j)}_{r,s}, \nu I)$. Here ν and the block size are tuning parameters, while I is an identity matrix of the appropriate size. The acceptance probability of updating $\boldsymbol{\alpha}^{(j)}_{r,s}$ to $\boldsymbol{\alpha}^{*}_{r,s}$ is given by

$$\min\left\{1, \frac{f(\boldsymbol{\alpha}^{*}|\mathbf{y}, \boldsymbol{\beta}^{(j)})}{f(\boldsymbol{\alpha}^{(j)}|\mathbf{y}, \boldsymbol{\beta}^{(j)})}\right\},$$

where corresponding elements of $(\boldsymbol{\alpha}^{(j)}_{r,s}, \boldsymbol{\alpha}^{*}_{r,s})$ outside the current block are identical.

4.4.4 Sampling from $f(\Sigma_\beta|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \Phi)$

The full conditional of Σ_β is made up of n Gaussian distributions and a conjugate Inverse-Wishart$(n_\Sigma, S_\Sigma^{-1})$ prior, which results in an

$$\text{Inverse-Wishart}\left(n_\Sigma + n, \left[S_\Sigma + \sum_{i=1}^{n}(\beta_i - \Phi_1\beta_{i-1} - \ldots - \Phi_p\beta_{i-p})(\beta_i - \Phi_1\beta_{i-1} - \ldots - \Phi_p\beta_{i-p})^{\mathrm{T}}\right]^{-1}\right)$$

posterior distribution. However the models applied in the case study are based on univariate autoregressive processes, meaning that the conjugate prior simplifies to an inverse-gamma(e, f) distribution, resulting in an

$$\text{Inverse-Gamma}\left(e + \frac{n}{2}, f + \frac{1}{2}\sum_{i=1}^{n}(\beta_i - \phi_1\beta_{i-1} - \ldots - \phi_p\beta_{i-p})^2\right)$$

posterior. If a non-informative prior is required an inverse-gamma(ϵ, ϵ) distribution with small ϵ is typically used. However as previously discussed an informative prior may be required to smooth the autoregressive process, and representing such beliefs using a member of the inverse-gamma family is not straightforward. This parameter is likely to be close to zero (to ensure the evolution is smooth), so a simple informative prior is a Gaussian distribution with zero mean that is truncated to be positive. The informativeness of this prior is controlled by its variance, with smaller values resulting in a more informative distribution. If this prior is used the full conditional distribution is non-standard, and can be sampled from using a Metropolis-Hastings step with a random walk proposal.

4.4.5 Sampling from $f(\Phi|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_\beta)$

The full conditional of Φ consists of n Gaussian distributions for $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ and a prior $f(\Phi)$, the latter of which depends on the form of the autoregressive process. The full conditional of Φ also depends on the type of autoregressive process, and two examples are univariate AR(1) ($\beta_i \sim \mathrm{N}(\phi_1\beta_{i-1}, \sigma_\beta^2)$), and AR(2) ($\beta_i \sim \mathrm{N}(\phi_1\beta_{i-1} + \phi_2\beta_{i-2}, \sigma_\beta^2)$) processes. In either case, assigning (ϕ_1) or (ϕ_1, ϕ_2) flat priors results in a Gaussian full conditional distribution, and details for the AR(1) case are given below.

AR(1) full conditional

$$\begin{aligned} f(\phi_1|\boldsymbol{\beta}, \sigma_\beta^2) &\propto \prod_{i=1}^{n}\mathrm{N}(\beta_i|\phi_1\beta_{i-1}, \sigma_\beta^2), \\ &\propto \mathrm{N}(\phi_1|\mu, \tau^2), \end{aligned}$$

where

$$\mu = \frac{\sum_{i=1}^{n} \beta_i \beta_{i-1}}{\sum_{i=1}^{n} \beta_{i-1}^2},$$
$$\tau^2 = \frac{\sigma_\beta^2}{\sum_{i=1}^{n} \beta_{i-1}^2}.$$

In a univariate AR(2) process (ϕ_1, ϕ_2) have a similar joint full conditional distribution. Alternatively if Φ only contains known constants, as in a random walk or local linear trend, then no simulation is required.

4.5 Case study

This case study analyses data from Greater London, to assess the efficacy of using dynamic generalised linear models in air pollution and health studies. The first subsection describes the data that are used in this case study, the second discusses the choice of statistical models, while the third presents the results.

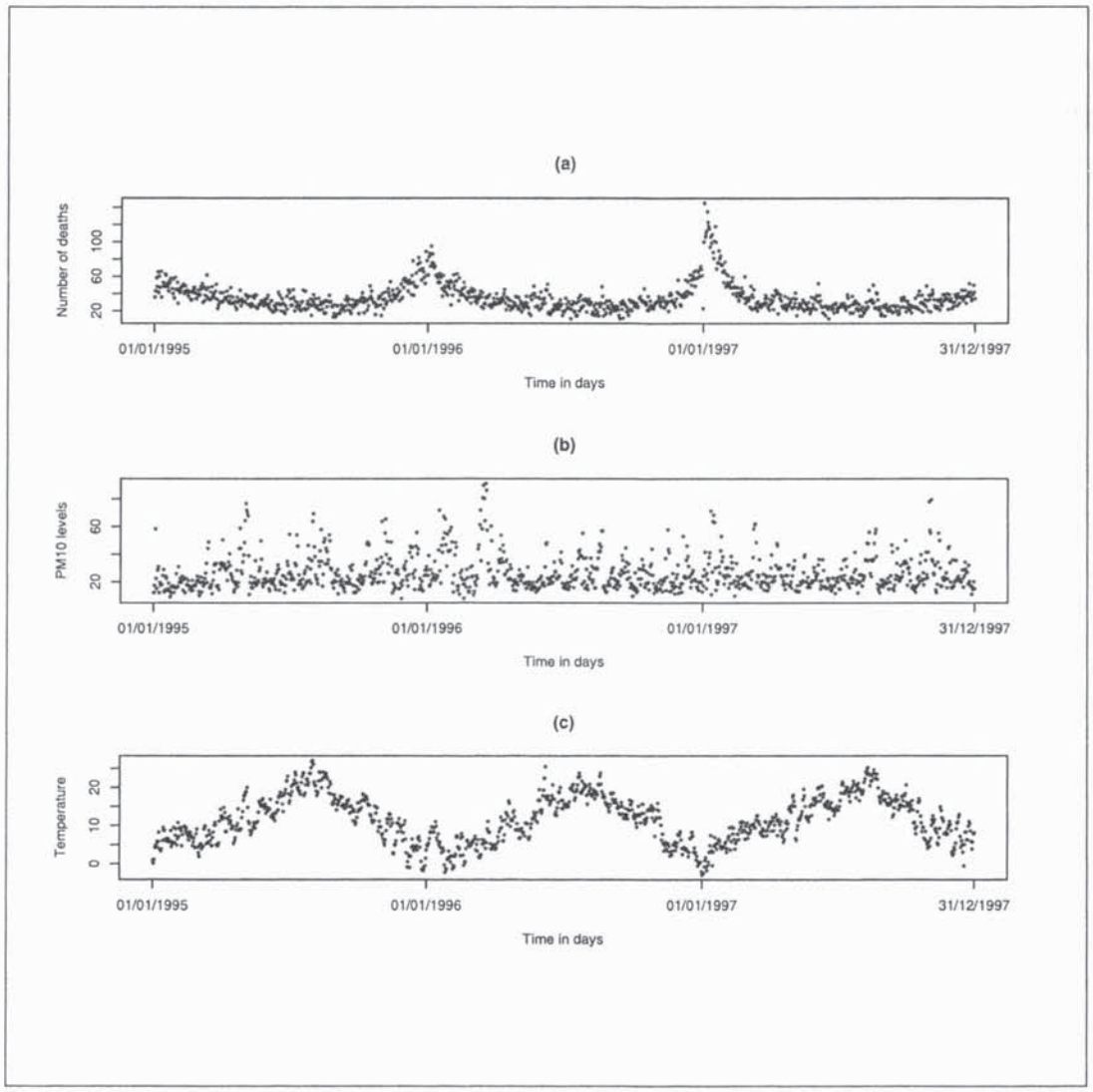
4.5.1 Description of the data

The data used in this case study relate to daily observations from the Greater London area during the period 1st January 1995 until 31st December 1997. The health data comprise daily counts of respiratory mortality drawn from the population living within Greater London, and are shown in Figure 4-1. A strong seasonal pattern is evident, with a large increase in the number of deaths during the winter of 1996/1997. The cause of this peak is unknown, and research has shown no influenza epidemic during this time which has previously been associated with large increases of this type (Griffin and Neuzil (2002)). The air pollution data comprise particulate matter levels measured as PM_{10}, which are shown in Figure 4-1. These data are collected from eleven monitoring sites located throughout Greater London, and a daily average is calculated to estimate pollution exposure. This spatial average has also been used by Katsouyanni et al. (1996) and Samet et al. (2000), and is likely to introduce minimal exposure error because PM_{10} levels in London between 1994 and 1997 exhibit little spatial variation (Shaddick and Wakefield (2002)). Meterological data are also available for Greater London, and include indices of temperature, rainfall, wind speed and sunshine. These data are measured at Heathrow airport, and daily mean temperature (measured in Celsius ^{0}C) is shown in Figure 4-1.

4.5.2 Description of the statistical models

Dynamic generalised linear models allow the standard approaches to analysing air pollution and health data to be extended in two ways: (i) modelling any long-term trends, over-dispersion and temporal correlation present in the health data with an autoregressive

Figure 4-1: Respiratory mortality (a), particulate matter (PM_{10}, b) and temperature (c) data from Greater London between 1995 and 1997.



process; (ii) allowing the relationship between air pollution and health to evolve over time. These extensions are investigated by applying eight models to the Greater London data, a summary of which is given in Table 4.1. The general form of all eight models is given by

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= \text{PM}_{10_{i-1}}\gamma_i + \beta_i + S(\text{temperature}_i|3, \boldsymbol{\alpha}_3), \\ \boldsymbol{\alpha} &\sim \text{N}(\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \end{aligned} \tag{4.6}$$

where β_i is the trend component and γ_i represents the effect of PM_{10} on day i. The trend component is represented by one of four sub-models (denoted by (a) - (d) below) including a natural cubic spline of calendar time or one of the autoregressive processes in equation (4.2).

(a) Natural cubic splines

$$\beta_i = \alpha_1 + S(i|27, \boldsymbol{\alpha}_2)$$

(b) First order random walk

$$\begin{aligned} &\beta_i \sim \text{N}(\beta_{i-1}, \tau^2) \\ &\beta_0 \sim \text{N}(3.5, 10) \\ &\tau^2 \sim \text{N}(0, g_2)_{I[\tau^2>0]} \end{aligned}$$

(c) Second order random walk

$$\begin{aligned} &\beta_i \sim \text{N}(2\beta_{i-1} - \beta_{i-2}, \tau^2) \\ &\beta_{-1}, \beta_0 \sim \text{N}(3.5, 10) \\ &\tau^2 \sim \text{N}(0, g_3)_{I[\tau^2>0]} \end{aligned}$$

(d) Local linear trend

$$\begin{aligned} &\beta_i \sim \text{N}(\beta_{i-1} + \delta_{i-1}, \tau^2) \\ &\delta_i \sim \text{N}(\delta_{i-1}, \psi^2) \\ &\beta_0 \sim \text{N}(3.5, 10) \\ &\delta_0 \sim \text{N}(0, 10) \\ &\tau^2 \sim \text{N}(0, g_4)_{I[\tau^2>0]} \\ &\psi^2 \sim \text{N}(0, g_5)_{I[\psi^2>0]} \end{aligned} \tag{4.7}$$

The relationship between air pollution and mortality is represented by (i) or (ii) below, which force it to be constant or allow it to evolve over time.

Table 4.1: Summary of the eight models. The base model is given by equation (4.6).

Model	Trend model β_i	Air pollution effect γ_i
1	(a) - splines	(i) - constant
2	(a) - splines	(ii) - random walk
3	(b) - first order random walk	(i) - constant
4	(b) - first order random walk	(ii) - random walk
5	(c) - second order random walk	(i) - constant
6	(c) - second order random walk	(ii) - random walk
7	(d) - local linear trend	(i) - constant
8	(d) - local linear trend	(ii) - random walk

(i) Constant	(ii) Time-varying - first order random walk	
$\gamma_i = \gamma$	$\gamma_i \sim \mathrm{N}(\gamma_{i-1}, \sigma^2)$	(4.8)
	$\gamma_0 \sim \mathrm{N}(0, 10)$	
	$\sigma^2 \sim \mathrm{N}(0, g_1)_{I[\sigma^2>0]}$	

In the model description above $\mathrm{N}(0, g_1)_{I[\sigma^2>0]}$ denotes a truncated Gaussian distribution which forces σ^2 to be positive. The smooth functions are estimated with natural cubic splines, where *var* denotes the covariate and *df* is the degrees of freedom. The vector of static parameters ($\boldsymbol{\alpha}$) is different for each model, and includes the intercept term, the parameters that make up the natural cubic splines and the constant effect of air pollution. To compare the results with those presented by Chiogna and Gaetan (2002), each model is analysed using the Bayesian approach described here and the likelihood based alternative that they adopt. Likelihood based analysis is carried out using the iteratively re-weighted Kalman filter and smoother proposed by Fahrmeir and Wagenpfeil (1997), while the hyperparameters are estimated using Akaike Information Criterion (AIC). The remainder of this subsection describes the model building process, including the choice of trend models and air pollution components.

Model building strategy

The model building process began by removing the long-term trend, seasonal variation, over-dispersion and temporal correlation from the respiratory mortality data. These data exhibit a pronounced yearly cycle, which is jointly modelled by the trend component β_i and the temperature covariate. The relationship between temperature and mortality was investigated at a number of lags using different shaped relationships, and the fit to the data was assessed via the deviance information criterion (DIC, Spiegelalter et al. (2002)). This

resulted in a smooth function of the same days temperature with three degrees of freedom being used in the final models, because it has the lowest DIC and because temperature exhibits a U-shaped relationship with mortality. The smooth function is modelled with a natural cubic spline because it is fully parametric, making analysis within a Bayesian setting straightforward. The relationships between numerous additional covariates and mortality were also investigated, including the amount of rainfall, number of sunshine hours, average wind speed and a day of the week effect. However none of these exhibited any relationship with the mortality data, and were not used in the final analyses.

After the meteorological covariates had been chosen, the smoothing parameters for each trend model (see Table 4.1) were investigated. The smooth function of calendar time (trend component (a)) is modelled by a natural cubic spline, and it's smoothness is determined by its degrees of freedom. In common with the temperature covariate this was chosen by DIC and fixed prior to analysis, resulting in 27 degrees of freedom. To allow a fairer comparison with trends (b) to (d) the degrees of freedom should be estimated as part of the MCMC algorithm, but this makes the average trend impossible to estimate. As the smoothness of the spline is fixed, its parameters (part of α) are assigned a non-informative Gaussian prior with a diagonal variance matrix. In the likelihood implementation the degrees of freedom are estimated by AIC to be 27, which is identical to the Bayesian estimate.

The remaining trend models are autoregressive processes, whose smoothness is determined by the evolution variances (τ^2, ψ^2). Initially these were assigned non-informative inverse-gamma$(0.01, 0.01)$ priors, which resulted in an estimated trend that close to interpolates the mortality data (see Figure 4-2 which shows the estimate for the second order random walk model). This undesirable interpolation can be removed by assigning (τ^2, ψ^2) informative priors, which shrink their estimates towards zero resulting in a smoother trend. Determining an informative prior distribution from the inverse-gamma family is not straightforward, and instead prior beliefs are represented by a Gaussian distribution with mean zero that is truncated to be positive. This prior forces (τ^2, ψ^2) to be close to zero, with the level of informativeness controlled by the variance parameters g_2 (first order random walk), g_3 (second order random walk) and (g_4, g_5) (local linear trend). As these variances decrease the prior mass is forced closer to zero, resulting in a smoother estimate of β.

It seems likely that the trend in mortality will be similar on consecutive days, meaning that the autoregressive process should evolve smoothly over time. The trend is estimated on the linear predictor scale corresponding to the natural log of the data, which ranges between 2.5 and 4.5 daily deaths (between about 12 and 90 on the un-logged scale). On that scale a jump of 0.01 between consecutive days is approximately the largest difference that cannot be detected by the eye, resulting in a visually smooth trend. To relate

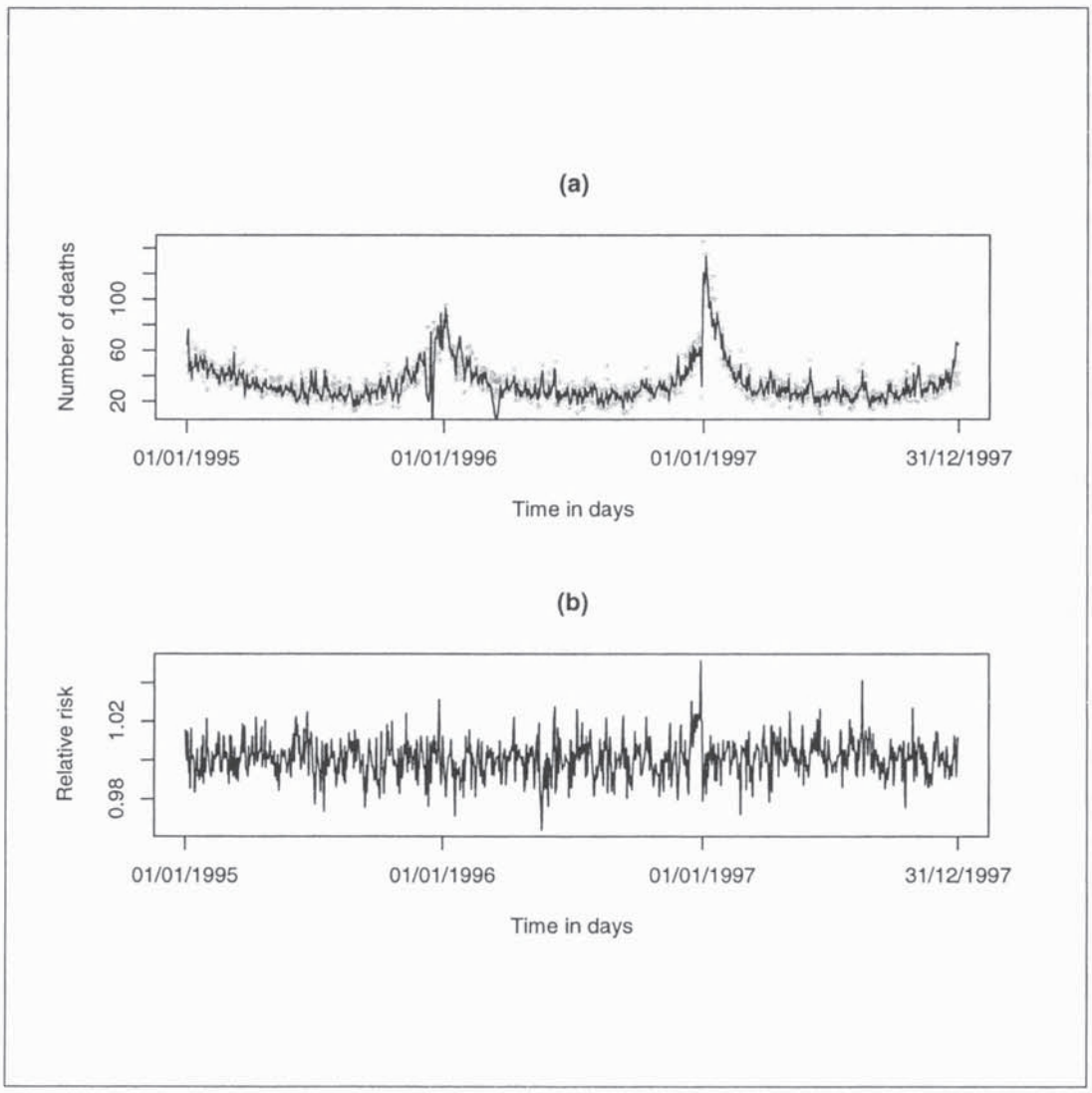
this to the choice of (g_2, g_3, g_4, g_5) each process was simulated with a variety of variances, and the average absolute difference between consecutive values was calculated. From this (g_2, g_3, g_4, g_5) were chosen so that 50% of the mass from $f(\tau^2)$ is below the threshold value giving average differences of 0.01, which resulted in $g_2 = 10^{-7}$, $g_3 = 10^{-14}$. The local linear trend model has two variance parameters that both need to be tightly controlled resulting in $g_4 = g_5 = 10^{-16}$. Sensitivity analyses were carried out for different values of (g_2, g_3, g_4, g_5), but it was found that larger values resulted in trends that were not visually smooth. In the likelihood implementation (τ^2, ψ^2) are chosen by optimising AIC. The priors for $(\beta_{-1}, \beta_0, \delta_0)$ are non-informative Gaussian distributions with means equal to zero (for the rate δ_0) and 3.5 ((β_{-1}, β_0), the average of logged mortality data from previous years) respectively.

After removing the influence of unmeasured risk factors the relationship between PM_{10} and mortality was investigated at a number of lags. A lag of one day is used in the final models because it has the minimum DIC and has been used in other recent studies (see for example Dominici et al. (2000), and Zhu et al. (2003)). Both constant and time-varying relationships between PM_{10} and mortality are investigated in this case study, with the latter modelled by a first order random walk which allows a comparison with the work of Chiogna and Gaetan (2002). Initially a non-informative inverse-gamma(0.01, 0.01) prior was specified for the random walk variance (denoted by σ^2), but as shown in Figure 4-2 the estimated time-varying relationship is contaminated with noise and an underlying trend cannot be seen. This relationship is likely to evolve smoothly over time, which can be enforced by assigning σ^2 an informative zero mean Gaussian prior that is truncated to be positive. The informativeness is controlled by the variance g_1, which is chosen using an identical approach to that described above. In this case the likely range of effects is -0.003 to 0.005, and the largest difference that is undetectable by the eye is around 0.00005 leading to $g_1 = 10^{-16}$. This value was investigated in a sensitivity analysis, which showed that γ_i evolved smoothly for σ^2 as large as 10^{-10}. As this is less informative than 10^{-16} it is used in the final models. In the likelihood implementation the variance parameter is estimated by optimising AIC.

4.5.3 Results

The models contain a large number of parameters, so to aid convergence the covariates (PM_{10} and the basis functions for the natural cubic splines of calendar time and temperature) are standardised to have a mean of zero and a standard deviation of one before inclusion in the model (and are subsequently back-transformed when obtaining results from the posterior distribution). The Markov chains are burnt in for 40,000 iterations, by which point convergence was assessed to have been reached using the methods of Gelman et al. (2003). At this point a further 100,000 iterations are simulated, which are

Figure 4-2: The estimated random walk processes using a non-informative prior for the evolution variance. Panel (a) depicts the estimated trend in daily mortality modelled as a second order random walk, while panel (b) depicts the time-varying relative risk between PM_{10} and mortality using a first order random walk.



thinned by 5 to reduce autocorrelation resulting in 20,000 samples from the joint posterior distribution.

(i) - Results for the four trend models β_i

In this case study long-term trends, seasonal variation, over-dispersion and temporal correlation are removed from the respiratory mortality data with one of four trend models: a natural cubic spline of calendar time (models 1 and 2); a first order random walk (models 3 and 4); a second order random walk (models 5 and 6); and a local linear trend model (models 7 and 8). To aid clarity in the following discussion these approaches are compared and contrasted assuming the relationship between PM_{10} and mortality is constant (using the odd numbered models). Figure 4-3 shows the mortality data from Greater London, together with the estimated trends from the Bayesian (solid lines) and likelihood (dotted lines) analyses. Panel (a) depicts the estimate from the natural cubic spline of calendar time, panel (b) relates to the first order random walk, panel (c) to the second order random walk, and panel (d) to the local linear trend model. All four models capture the underlying trend in the mortality series well, and the Bayesian and likelihood estimates are very similar. The only major differences are in the winters of 1996 and 1997, where the respiratory mortality data has yearly peaks. For each trend model the Bayesian analysis captures the height of these peaks better than its likelihood counterpart, while the second order random walk outperforms the three alternatives. For example, in the winter of 1997 the maximum number of deaths on a single day is 145, and the Bayesian estimates of this peak are, (a) 98.4, (b) 92.2, (c) 107.6, (d) 101.5, while the corresponding likelihood values are, (a) 69.4, (b) 79.1, (c) 85.9, (d) 85.1. These figures show that the second order random walk is the most adept at modelling these peaks, while the local linear trend model outperforms both the natural cubic spline and the first order random walk.

All eight estimates have the same visual smoothness, and a summary of their smoothing parameters is given in Table 4.2. For the natural cubic spline model the degrees of freedom are identical in the Bayesian and likelihood analyses, which is probably because they are estimated by minimising similar data driven criteria (DIC and GCV respectively). However, for the other trend models the likelihood estimates of the smoothing parameters are significantly larger than their Bayesian counterparts, without the corresponding trends being less smooth. This is unexpected, and is most likely caused by differences in the techniques used to estimate the autoregressive processes, a point which is taken up in the discussion. The trend components should remove any temporal correlation in the respiratory mortality series, and their effectiveness in this respect can be assessed by examining the residuals. However in a Bayesian setting residuals are not well defined (see Pettit (1986)), because there is no natural point estimate for the parameters. Alternatively a 'residual distribution' can be generated for each y_i by simulation, which for standardised

Table 4.2: Summary of the smoothing parameters.

Model	Parameter	Bayesian			Likelihood
		2.5%	median	97.5%	
1	k	-	27	-	27
2	k	-	27	-	27
	σ^2	9.91×10^{-8}	1.10×10^{-7}	1.21×10^{-7}	0
3	τ^2	0.00018	0.00019	0.00021	0.373
4	τ^2	0.00018	0.00019	0.00021	0.373
	σ^2	1.01×10^{-7}	1.10×10^{-7}	1.20×10^{-7}	0
5	τ^2	2.20×10^{-6}	3.56×10^{-6}	5.78×10^{-6}	0.004
6	τ^2	2.27×10^{-6}	3.67×10^{-6}	5.95×10^{-6}	0.004
	σ^2	1.19×10^{-7}	8.25×10^{-7}	1.66×10^{-6}	0
7	τ^2	1.61×10^{-7}	1.68×10^{-7}	1.76×10^{-7}	10^{-7}
	ψ^2	3.00×10^{-6}	3.09×10^{-6}	3.18×10^{-6}	0.003
8	τ^2	1.58×10^{-7}	1.67×10^{-7}	1.77×10^{-7}	10^{-7}
	ψ^2	3.05×10^{-6}	3.17×10^{-6}	3.25×10^{-6}	0.003
	σ^2	7.43×10^{-7}	2.64×10^{-6}	5.44×10^{-6}	0

residuals is given by

$$r_i^{(j)} = \frac{y_i - \mathbb{E}[y_i|\boldsymbol{\theta}^{(j)}]}{\sqrt{\mathbb{E}[y_i|\boldsymbol{\theta}^{(j)}]}},$$

where $\boldsymbol{\theta}^{(j)}$ is the jth sample from the joint posterior distribution. This distribution takes into account the uncertainty in $\boldsymbol{\theta}$, and residuals based on point estimates are approximations to this distribution. Figure 4-4 shows the autocorrelation function of an approximation to this residual distribution that is based on posterior medians. The second order random walk again outperforms the other approaches, exhibiting little or no correlation in the approximate standardised residuals. In contrast the natural cubic spline has the worst performance of the four trend components, exhibiting significant correlation at the first four lags. The remaining two models produce similar results, exhibiting significant correlation at the first lag only. The residuals from the likelihood analyses are shown in Figure 4-5 and a similar pattern is evident. A comparison of Figures 4-4 and 4-5 shows that the residuals from the likelihood analysis exhibit much larger correlation, suggesting that the Bayesian models outperformed their likelihood counterparts in all four cases.

(ii) - Results for the air pollution and mortality relationship γ_i

The models presented here allow the relationship between PM_{10} and mortality to evolve over time as a first order random walk, or be fixed at a constant value. In the graphs and tables that follow, these relationships are presented as relative risks for an increase of

Figure 4-3: Observed mortality counts and the estimated trends from the four models: (a) natural cubic spline, (b) first order random walk, (c) second order random walk, (d) local linear trend. Bayesian and likelihood estimates are represented by solid and dotted lines respectively.

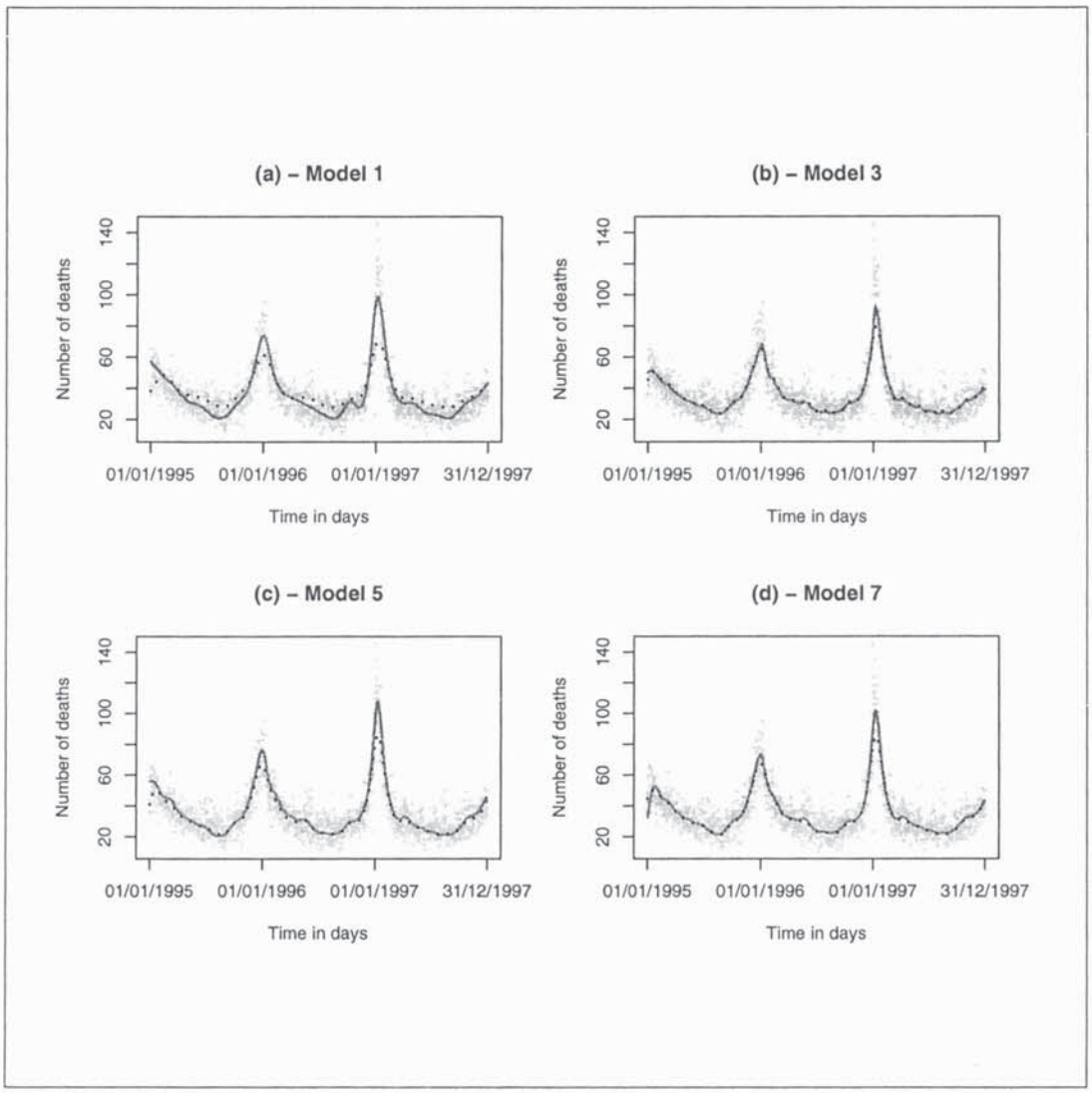


Figure 4-4: Autocorrelation function of the Bayesian residuals from using each trend model: (a) natural cubic spline, (b) first order random walk, (c) second order random walk, (d) local linear trend.

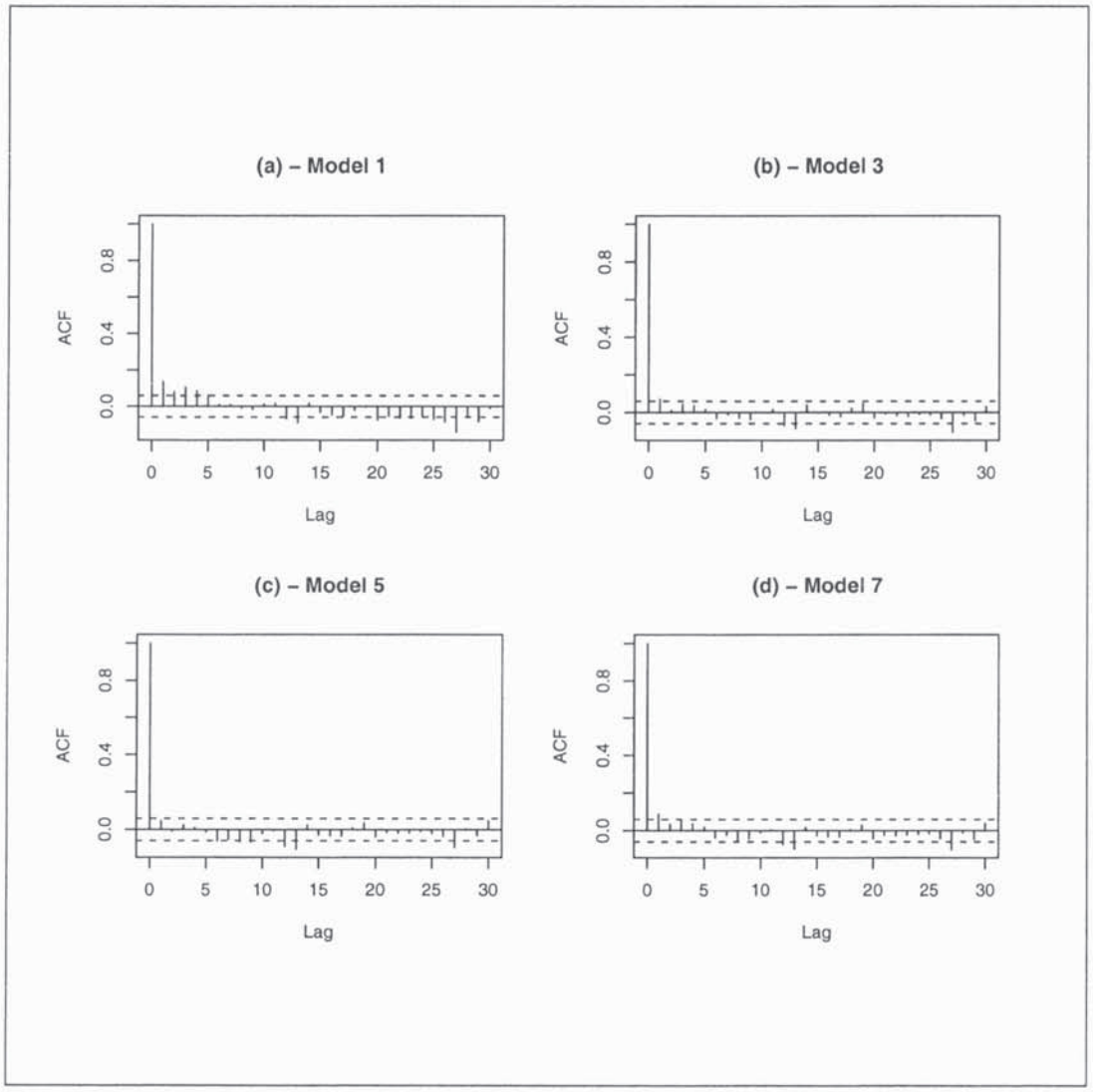


Figure 4-5: Autocorrelation function of the likelihood residuals from using each trend model: (a) natural cubic spline, (b) first order random walk, (c) second order random walk, (d) local linear trend.

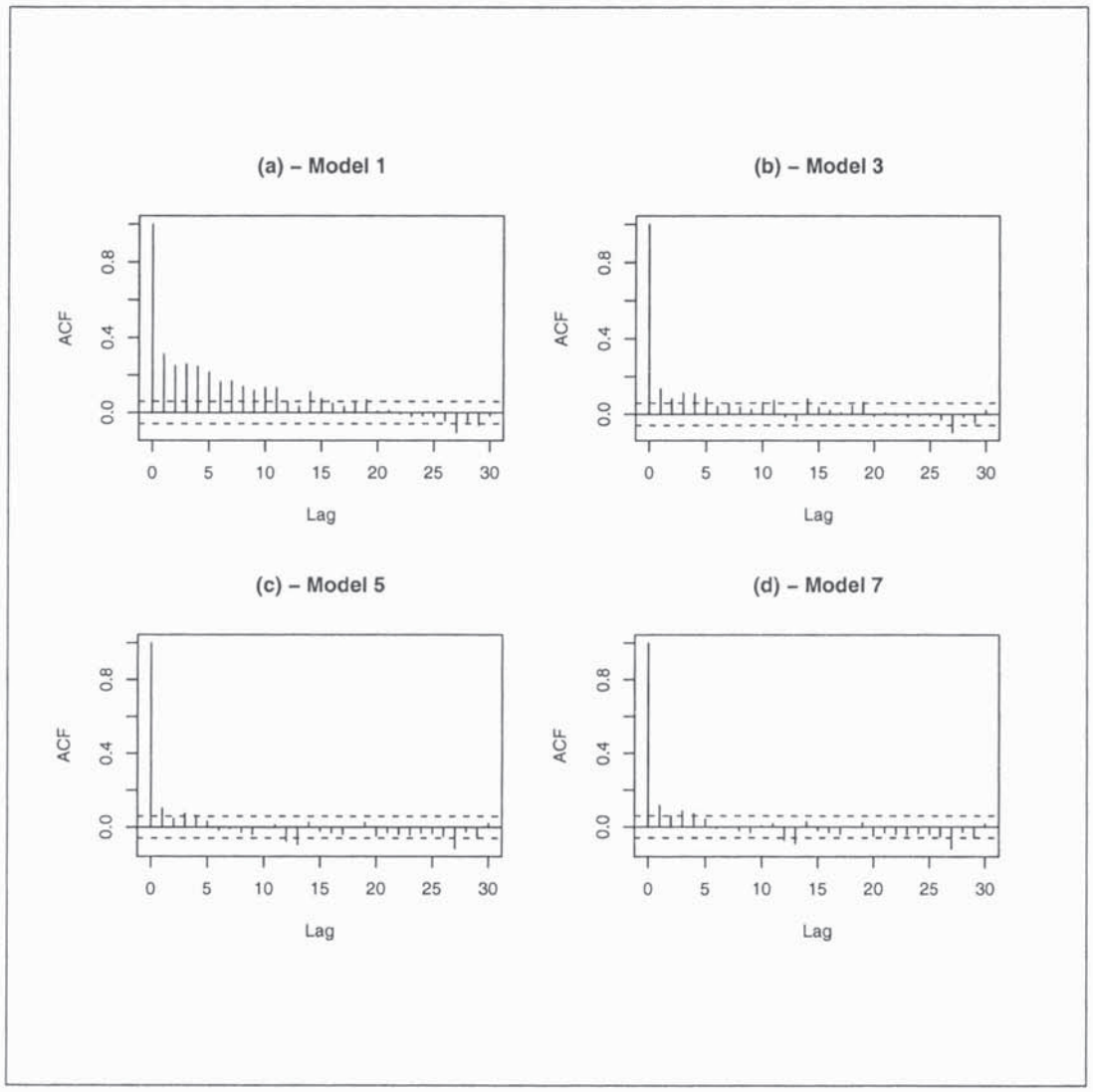


Table 4.3: Relative risks for an increase of $10\mu g/m^3$ in PM_{10} with associated 95% credible or confidence intervals.

Model	Bayesian	Likelihood
1	1.007 (0.998 , 1.016)	1.015 (1.010 , 1.020)
3	1.011 (1.002 , 1.020)	1.014 (1.008 , 1.019)
5	1.008 (0.999 , 1.017)	1.014 (1.008 , 1.019)
7	1.009 (1.000 , 1.019)	1.013 (1.007 , 1.018)

$10\mu g/m^3$ in PM_{10}.

(a) - Constant relationship

Table 4.3 shows the estimated relative risks from models 1, 3, 5 and 7, which force the relationship between PM_{10} and mortality to be constant. All eight Bayesian and likelihood estimates are very similar (range from 1.007 to 1.015), suggesting that the method of analysis and the choice of trend component do not substantially affect the estimated health risk. The estimates from the likelihood analyses are always larger than those from the corresponding Bayesian model, although the differences are not large. The Bayesian credible intervals are wider than their likelihood counterparts while few of the intervals contain one, suggesting that exposure to PM_{10} has a statistically significant effect on mortality.

(b) - Time-varying relationship

In the likelihood analyses the estimated variance parameters are all zero, forcing the time-varying relationship to be constant. However the Bayesian estimates are greater than zero, and the time-varying relationships are shown in Figure 4-6. The temporal evolution for each of these estimates is smooth, which is a result of the informative prior placed on σ^2. All four estimates are very similar, suggesting that the choice of trend model does not affect the substantive conclusions. The relationships exhibit a slowly increasing long-term trend, which has ranges of: (a) 1.005 to 1.015, (b) 1.007 to 1.014, (c) 1.002 to 1.019, (d) 0.999 to 1.024. In particular the time-varying relationships do not exhibit a seasonal pattern, suggesting that the model of Peng et al. (2005) would be inappropriate for these data. The 95% credible intervals for panels (a) and (b) (models 2 and 4 respectively) are of a similar size, while the remaining two exhibit substantial additional variation, especially in panel (d). This additional variation is not supported by the same pattern in the credible

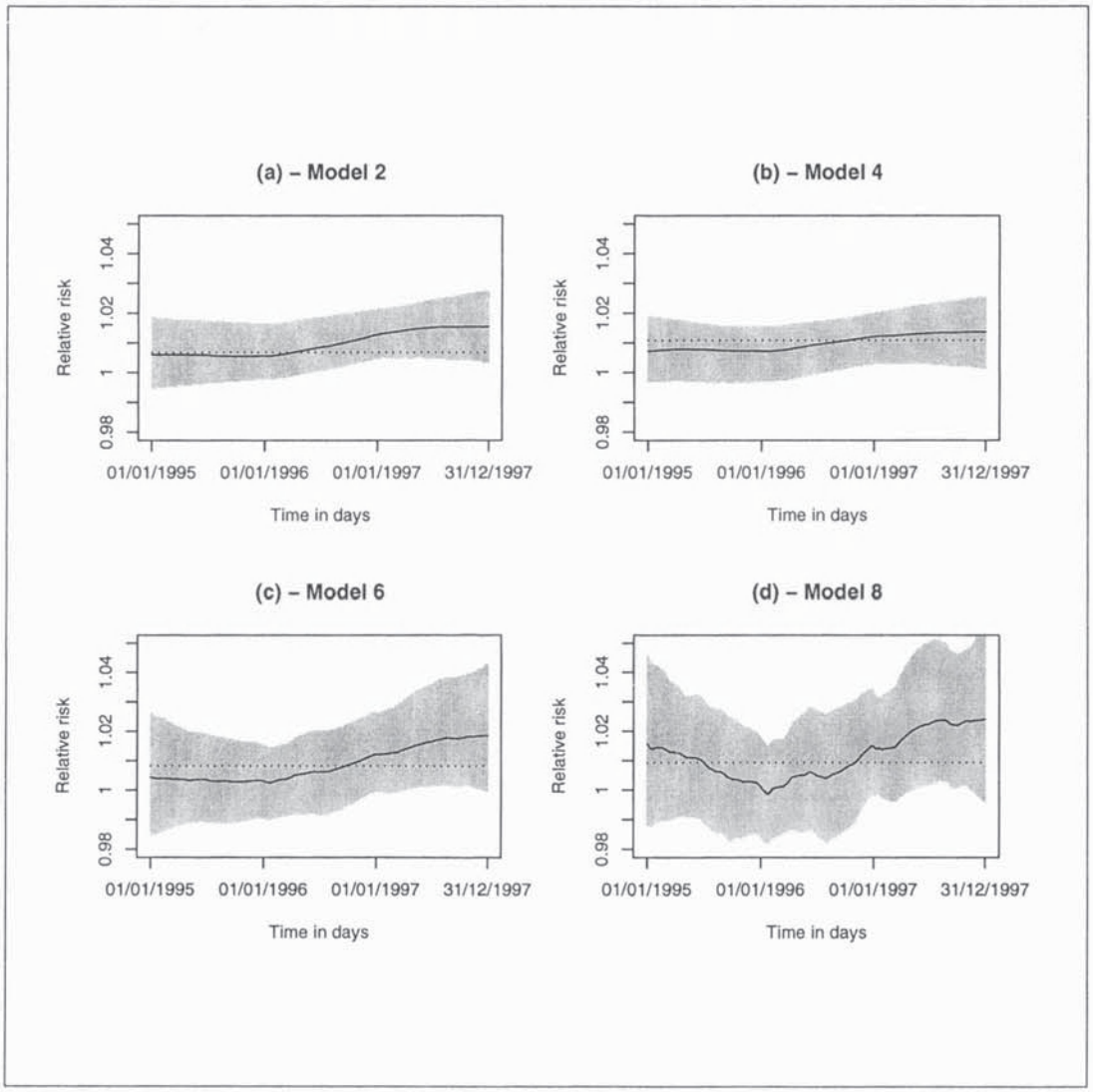
intervals for the constant effects, a point which is taken up in the discussion. However the width of these intervals suggest that a constant effect of PM_{10} cannot be ruled out.

4.6 Discussion

This chapter has investigated the efficacy of dynamic generalised linear models for analysing air pollution and health data, with particular interest in two differences between this dynamic framework and the standard approaches adopted in the related literature. The dynamic nature of the models proposed here allow the pollution-mortality relationship to change over time, whilst any long-term trend, seasonal variation, over-dispersion and temporal correlation can be modelled with an autoregressive process. The results from the four trend models lead to two main conclusions. Firstly although all models capture the underlying level of daily deaths well, the standard approach of using smooth functions is outperformed by the autoregressive processes. In particular the best of these is the second order random walk, because its residuals exhibit no correlation and the two winter peaks in daily mortality are well represented. In contrast the smooth function leaves significant correlation in the residuals, while the estimated peaks in mortality are poorly captured. The local linear trend model also outperforms the smooth function, while the first order random walk gives similar results. The poor performance of the smooth function most likely results from the way it is estimated, including the smoothing parameter and use of natural cubic splines. The degrees of freedom is estimated by DIC and fixed during the simulation, which is in contrast to the autoregressive processes whose smoothing parameters are estimated as part of the MCMC algorithm. Consequently the autoregressive trends incorporate the variation in their smoothing parameters, which is not the case for the smooth function and may account for the latter's poorer performance. Another possible cause is the use of natural cubic splines, which were used here because of their parametric make-up. However they are less flexible than non-parametric alternatives, and an interesting area of future research would be to compare the performances of the trend models used here against non-parametric smooth functions such as smoothing splines or LOESS smoothers.

Secondly the Bayesian approach produces superior results to those obtained from the likelihood analysis, both in terms of removing temporal correlation from the health data and its ability to capture winter peaks in mortality. The estimated smoothing parameters for both implementations of the natural cubic splines are obtained by optimising data driven criteria (DIC and AIC), and it is not surprising that the estimates are identical. However, for the autoregressive processes the Bayesian estimates are smaller than their likelihood counterparts, which is caused by the relative strengths of the truncated Gaussian prior and the penalty term in the AIC criteria. A sensitivity analysis shows that such a strong prior is required for these data, because using a non-informative prior for τ^2 results in an esti-

Figure 4-6: Estimated time-varying relationships (solid line) between PM_{10} and mortality from the Bayesian analyses. The dotted lines depict a constant association over time, while the grey shading represents 95% credible intervals. The four panels relate to the different trend models: (a) natural cubic spline, (b) first order random walk, (c) second order random walk, (d) local linear trend.



mated trend that interpolates the mortality series. This interpolation is probably caused by non-identifiability problems that arise from modelling the trend with a large number of parameters, and can be removed by imposing a more stringent autoregressive constraint. An initial comparison of the Bayesian and likelihood smoothing parameters shows that the latter are larger, and therefore expected to produce a trend exhibiting greater variability (thus modelling the peaks in the mortality data more accurately). However the opposite effect is observed, as the larger likelihood estimates result in trends which are less variable. This apparent anomaly is probably caused by the methods used to implement the autoregressive constraint, which in the Bayesian analysis is implemented via an autoregressive prior. In contrast the likelihood approach is based on the Kalman filter, which uses a two stage process to estimate the trend. Firstly $\mathbb{E}[\beta_i|y_1,\ldots,y_i]$ is estimated for all i, which are subsequently smoothed by estimating $\mathbb{E}[\beta_i|y_1,\ldots,y_n]$. The final likelihood estimates are based on these smoothed values, and it is this additional smoothing imposed by the Kalman filter that reduces the variability in the estimated trends over-smoothing the data in this case.

The Bayesian estimates of the pollution-mortality relationship exhibit a consistent long-term pattern regardless of the model, suggesting that this temporal variation should be investigated further. However no seasonal interaction is observed for these data, meaning that the model of Peng et al. (2005) would be too restrictive. The informative prior for σ^2 forces this relationship to evolve smoothly over time, while a sensitivity analysis showed that using a non-informative prior leads to the estimate being contaminated with unwanted noise. This noise is caused by non-identifiability problems that arise from using an excess number of parameters to model the time-varying relationship. In contrast an informative prior shrinks the evolution variance towards zero, which reduces the number of effective parameters and leads to a smoother estimate. However this is achieved at the expense of a very informative prior. The estimated time-varying relationship is not altered by the choice of trend model, although the credible intervals increase in width if a second order random walk or local linear trend are used. These two represent the most flexible trend models adopted in this case study, and their increased variation may cause slight non-identifiability or collinearity with the time-varying pollution-mortality relationship, reducing its precision. The Bayesian estimates of the pollution-mortality relationship exhibit greater curvature than their likelihood counterparts, which is probably caused by the methods used to estimate the respective smoothing parameters. The likelihood approach calculates the likelihood for a range of values of the smoothing parameter, and estimates σ^2 by optimising a data driven criterion. In contrast the Bayesian approach averages over the posterior for σ^2, which incorporates the possibility of no smoothing leading to an estimate which exhibits greater curvature.

Chapter 5

Investigating the time-varying relationship between air pollution and mortality

5.1 Introduction

In this chapter I use a general class of time-varying coefficient models (TVCM) to investigate whether the relationship between air pollution and mortality changes over time. This investigation extends the work presented in the previous chapter, where the temporal evolution was modelled as a first order random walk within a dynamic generalised linear model. This is a relatively new area of research, as few studies have estimated the temporal variation in the health risk associated with air pollution exposure. Those that have examined this possibility have either fixed the relationship to follow a simple parametric form or allowed it to evolve freely as an autoregressive process, neither of which maybe ideal. The time-varying coefficient model proposed here is an improvement on these approaches, because it combines the flexibility of the random walk with the smoothness of the simple parametric forms. The proposed model is implemented within both Bayesian and likelihood settings, and its efficacy is investigated using real and simulated data.

The remainder of this chapter is organised as follows. Section 5.2 reviews the background and motivation for this work, including a classification of the air pollution and health relationships that have been adopted in the related literature. Section 5.3 proposes a time-varying coefficient model for analysing air pollution and health data, which represents any temporal variation in this relationship as a flexible smooth function. Section 5.4 presents Bayesian and likelihood implementations of this model, the former being based on MCMC simulation while the latter uses an iterative maximisation algorithm. Section 5.5 presents a simulation study, which examines the ability of the proposed model to estimate different

shaped time-varying relationships. Section 5.6 examines the relationships between air pollution and mortality across four U.S. cities, by using the proposed model and a series of simpler alternatives. Finally section 5.7 presents a concluding discussion.

5.2 Background and motivation

The shape of the air pollution and mortality relationship has been the subject of much recent research, and is typically estimated within a Poisson regression model. A general form of this model is given by

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= w_i\gamma(\xi_i) + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\alpha}, \end{aligned} \tag{5.1}$$

where w_i denotes air pollution exposure and $\gamma(\xi_i)$ represents its relationship with mortality or morbidity. Equation (5.1) is an example of a varying coefficient model (VCM, Hastie and Tibshirani (1993)), in which the relationship of interest ($\gamma(\xi_i)$) depends on a known covariate ξ_i, called an effect modifier. The majority of existing air pollution and health studies have represented this relationship as one of three special cases:

(i) constant relationship - set $\xi_i = 1$ resulting in $w_i\gamma(\xi_i) = w_i\gamma$;

(ii) dose-response relationship - set $\xi_i = w_i$ and $w_i = 1$ resulting in $w_i\gamma(\xi_i) = \gamma(w_i)$;

(iii) time-varying relationship - set $\xi_i = i$ resulting in $w_i\gamma(\xi_i) = w_i\gamma_i$.

The existing research on each of these alternatives is summarised below, while a new approach to representing the time-varying relationship is discussed in section 5.3.

5.2.1 Constant relationship

The standard approach to modelling air pollution and health data fixes their relationship to be constant, which is typically assumed for simplicity because the relationship can be summarised by a single value. However the true relationship is likely to depend on a set of risk factors, meaning that a constant association is overly restrictive.

5.2.2 Dose-response relationship

Dose-response relationships allow the association between air pollution exposure and mortality or morbidity to depend on the pollution level, and have become increasingly popular in recent years. One of the first dose-response curves was estimated by Schwartz (1994b), who represented it as a smooth function within a generalised additive model. Since then numerous researchers have investigated dose-response relationships, modelling them as a

linear threshold model (Cakmak et al. (1999)), a smooth function (Schwartz and Zanobetti (2000)) and a piecewise linear function (Roberts (2004)). Recently a comprehensive study of dose-response curves was conducted by Daniels et al. (2004), who analysed data from 100 U.S. cities as part of the National Morbidity, Mortality and Air Pollution Study. They found that the shape of the curves varied between cities, but reported a national curve that was linearly increasing and exhibited no threshold level. Such results are the primary motivation for estimating dose-response curves, because the health risk associated with exposure to low levels of pollution is of regulatory interest.

5.2.3 Time-varying relationship

The existing research about time-varying pollution-mortality relationships was outlined in chapter four within the context of dynamic models, and is repeated here for completeness. Only a small number of studies have investigated this possibility, and if temporal variation exists it is likely to be seasonal or exhibit a long-term trend. A seasonal relationship may be caused by an interaction with temperature or another pollutant exhibiting a seasonal pattern, while a long-term trend may result from slow changes in the composition of individual pollutants or the size and structure of the population at risk. In addition to the case study described in the previous chapter, only Moolgavkar et al. (1995), Chiogna and Gaetan (2002) and Peng et al. (2005) have investigated time-varying relationships of this type. These studies model the relationship as:

$$\begin{aligned}
\text{Moolgavkar et al - } \gamma_i &= \begin{cases} \theta_1 & \text{for spring days} \\ \theta_2 & \text{for summer days} \\ \theta_3 & \text{for autumn days} \\ \theta_4 & \text{for winter days} \end{cases}, \\
\text{Chiogna et al - } \gamma_i &\sim \mathrm{N}(\gamma_{i-1}, \sigma^2), \\
\text{Peng et al - } \gamma_i &= \theta_1 + \theta_2 \sin(2\pi i/365) + \theta_3 \cos(2\pi i/365).
\end{aligned} \tag{5.2}$$

The relationships adopted by Moolgavkar et al. (1995) and Peng et al. (2005) are seasonal parametric forms, which are overly restrictive because they cannot exhibit non-seasonal variation. In contrast the first order random walk adopted by Chiogna and Gaetan (2002) does not fix the shape of the time-varying relationship, allowing it to be estimated from the data which results in a more realistic model. However as shown in the previous chapter (see Figure 4-2) a random walk may not evolve smoothly over time, meaning that the underlying shape may be hidden by unwanted noise. This noise is most likely due to non-identifiability problems that result from using a large number of parameters to estimate this relationship, and a more parsimonious representation is desirable. In this chapter I investigate the efficacy of using a smooth function to represent this relationship,

within the general setting of a time-varying coefficient model. A smooth function is an improvement on the previous approaches listed in equation (5.2), because the estimate will change smoothly over time without having a predetermined temporal shape. I assess the efficacy of using smooth functions for this purpose by answering two important questions: (i) can a smooth function accurately estimate different shaped time-varying relationships; (ii) how does its performance compare with previously used alternatives. In addition to assessing the usefulness of smooth functions in this setting, this chapter aims to investigate the existence of temporal variation in pollution-mortality relationships from real data.

5.3 Time-varying coefficient models

The time-varying coefficient model proposed here represents temporal variation in the pollution-health relationship as a flexible smooth function, and is given by

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= w_i\gamma_i + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\alpha}, \\ \gamma_i &= S(i|\boldsymbol{\theta}), \end{aligned} \tag{5.3}$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)_{n\times 1}$ denotes the complete relationship for all n days. The smooth function is represented here by a parametric regression spline because it is straightforward to implement in a Bayesian analysis. An alternative that is not followed up here is to represent $S(i|\boldsymbol{\theta})$ with a non-parametric smooth function, and estimate $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ using the methods of Lin and Zhang (1999). The regression spline represents the pollution-health relationship by $\boldsymbol{\gamma} = B\boldsymbol{\theta}$, where $B = (\mathbf{B}_1^{\mathrm{T}}, \ldots, \mathbf{B}_n^{\mathrm{T}})$ is an $n \times p$ matrix of known basis functions while $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ is a $p \times 1$ vector of unknown parameters. The relationship on day i is given by $S(i|\boldsymbol{\theta}) = \mathbf{B}_i^{\mathrm{T}}\boldsymbol{\theta}$, and can be incorporated into a generalised linear model structure. The basis functions can be B-splines or truncated polynomials, and the former are used here because they are numerically stable. The particular class of regression splines chosen to model $S(i|\boldsymbol{\theta})$ are natural cubic splines, because they are visually smooth and linear beyond the end knots which prevents erratic tail behaviour. The smoothness of the spline can be controlled in a variety of ways, and a penalised approach is adopted here because it is determined by a single parameter. An alternative is to determine the smoothness by choosing the number and locations of a set of knots, for which prior specification is unclear. The penalised approach represents the spline with an overly large number of basis functions, which allows it to exhibit substantial variation. This variation is reduced by imposing a smoothing penalty on $\boldsymbol{\theta}$, which lowers the effective number of parameters leading to a less rough estimate. A variety of penalties have been proposed for this problem, and the approach adopted here is that of Eilers and Marx (1996) who penalise kth order differences in $\boldsymbol{\theta}$. This penalty is adopted because it is straightforward

to implement within both Bayesian and likelihood settings, and is given by

$$\sum_{l=k+1}^{p} (\Delta^k \theta_l)^2,$$

where Δ denotes the difference operator ($\Delta^1 \theta_l = \theta_l - \theta_{l-1}$). Second order differences are used here, meaning that the penalty term is given by

$$\sum_{l=3}^{p} (\theta_l - 2\theta_{l-1} + \theta_{l-2})^2 = \theta^{\mathrm{T}} D \theta, \tag{5.4}$$

where

$$D = \begin{pmatrix}
1 & -2 & 1 & & & & & & \\
-2 & 5 & -4 & 1 & & & & & \\
1 & -4 & 6 & -4 & 1 & & & & \\
 & 1 & -4 & 6 & -4 & 1 & & & \\
 & & \vdots & \vdots & \vdots & \vdots & \vdots & & \\
 & & & 1 & -4 & 6 & -4 & 1 & \\
 & & & & 1 & -4 & 6 & -4 & 1 \\
 & & & & & 1 & -4 & 5 & -2 \\
 & & & & & & 1 & -2 & 1
\end{pmatrix}. \tag{5.5}$$

The likelihood implementation of the time-varying coefficient model is given by equations (5.3) and (5.4), where the latter is multiplied by a smoothing parameter λ in the estimation stage. Difference penalties were implemented in a Bayesian setting by Lang and Brezger (2004), who assigned $\boldsymbol{\theta}$ a kth order random walk prior. For the second order penalty adopted here the prior is given by

$$\theta_l \sim \mathrm{N}(2\theta_{l-1} - \theta_{l-2}, \sigma^2) \quad \text{for } l = 3, \ldots, p,$$

which effectively replaces (5.4) with its stochastic analogue. For convenience the first two parameters are assigned non-informative priors. The Bayesian time-varying coefficient model proposed here is given by

$$
\begin{aligned}
y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\
\ln(\mu_i) &= w_i \gamma_i + \mathbf{z}_i^{\mathrm{T}} \boldsymbol{\alpha}, \\
\gamma_i &= \mathbf{B}_i^{\mathrm{T}} \boldsymbol{\theta}, \\
\theta_l &\sim \text{N}(2\theta_{l-1} - \theta_{l-2}, \sigma^2) \quad \text{for } l = 3, \ldots, p, \\
\boldsymbol{\alpha} &\sim \text{N}(\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \\
\sigma^2 &\sim \text{inverse-gamma}(e, f),
\end{aligned} \tag{5.6}
$$

where σ^2 controls the smoothness of any temporal variation in the pollution-health relationship. As σ^2 increases this relationship exhibits greater variation, because the random walk constraint becomes weaker and allows the individual elements of $\boldsymbol{\theta}$ to be more dispersed. For convenience $(\sigma^2, \boldsymbol{\alpha})$ are assigned conjugate inverse-gamma and multivariate Gaussian priors respectively.

5.4 Estimation

This section describes Bayesian and likelihood estimation for the time-varying coefficient model proposed in the previous section. Bayesian estimation uses MCMC simulation while the likelihood implementation is achieved via an iterative maximisation algorithm.

5.4.1 Bayesian estimation

The Bayesian model is that of equation (5.6), for which the joint posterior distribution is given by

$$
\begin{aligned}
f(\boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\alpha}) f(\boldsymbol{\theta}|\sigma^2) f(\sigma^2) f(\boldsymbol{\alpha}), \\
&= \prod_{i=1}^{n} \text{Poisson}(y_i|\boldsymbol{\theta}, \boldsymbol{\alpha}) \prod_{l=3}^{p} \text{N}(\theta_l | 2\theta_{l-1} - \theta_{l-2}, \sigma^2) \\
&\times \text{Inverse-Gamma}(\sigma^2|e, f) \text{N}(\boldsymbol{\alpha}|\boldsymbol{\mu}_\alpha, \Sigma_\alpha),
\end{aligned}
$$

with flat priors assigned to (θ_1, θ_2). Simulation for this type of regression problem has been developed by Fahrmeir and Lang (2001), and the algorithm described below is based on their work. The parameters are updated using a block Metropolis-Hastings algorithm, in which starting values $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\alpha}^{(0)}, \sigma^{2(0)})$ are generated from over-dispersed versions of the priors (for example t-distributions replace Gaussian distributions). The algorithm alternately samples from the full conditional distributions of the following blocks.

(a) Penalised spline parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$.

Sampled using a Metropolis-Hastings algorithm with the conditional prior proposal method of Knorr-Held (1999).

(b) **Covariate parameters** $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$.
Sampled using a Metropolis-Hastings algorithm with a random walk proposal.

(c) **Smoothing parameter** σ^2.
Sampled using a Gibbs algorithm from its inverse-gamma distribution.

Specific details are given below.

(a) sampling from $f(\theta|\alpha, \sigma^2, \mathbf{y})$

The full conditional of θ (given below) is the product of n Poisson observations and a Gaussian second order random walk prior.

$$f(\theta|\alpha, \sigma^2, \mathbf{y}) \propto \prod_{i=1}^{n} \text{Poisson}(y_i|\theta, \alpha) \prod_{l=3}^{p} \text{N}(\theta_l|2\theta_{l-1} - \theta_{l-2}, \sigma^2)$$

This full conditional distribution is non-standard, meaning that a Metropolis-Hastings algorithm is typically used to update θ. It is nearly identical to that of the dynamic parameters in chapter four, and the same simulation scheme is used here. The proposal distribution is that suggested by Knorr-Held (1999, for details see chapter 4.4), which is also the approach adopted by Fahrmeir and Lang (2001).

(b) sampling from $f(\alpha|\theta, \sigma^2, \mathbf{y})$

The full conditional of α is also non-standard, because it is a product of n Poisson observations and a multivariate Gaussian prior. It is identical to the full conditional of the static parameters in chapter four (for details see chapter 4.4), and is updated using the same Metropolis-Hastings random walk algorithm.

(c) sampling from $f(\sigma^2|\theta, \alpha, \mathbf{y})$

The full conditional of σ^2 comprises $p-2$ Gaussian distributions and an inverse-gamma(e, f) prior, resulting in an

$$\text{Inverse-Gamma}\left(e + \frac{p-2}{2}, f + \frac{1}{2}\theta^{\text{T}} D \theta\right)$$

posterior distribution. The penalty matrix D is given by equation (5.5), and σ^2 is an important parameter in the model because it determines the amount of temporal variation in γ. The time-varying pollution-mortality relationship is likely to change smoothly over time, meaning that σ^2 may be very small and have a posterior distribution with significant mass near zero. In this scenario supposedly non-informative inverse-gamma(ϵ, ϵ)

priors (where $\epsilon = 1, 0.1, 0.01$) are of little use, because they give little or no prior weight to values very close to zero. Gelman (2006) has investigated prior choice in this situation, and suggests using a non-informative prior on the standard deviation scale, $f(\sigma) \propto 1$ (equivalently $f(\sigma^2) \propto (\sigma^2)^{-1/2}$). This prior leads to an Inverse-Gamma $\left(\frac{p-3}{2}, \frac{1}{2}\theta^{\mathrm{T}} D\theta\right)$ full conditional distribution, which effectively sets $e = -1/2$ and $f = 0$. A further alternative is to adopt the approach taken in chapter four and assign σ^2 a truncated Gaussian prior with zero mean. However in this setting where an informative prior is not required, the complexity introduced by its non-conjugacy seems redundant.

5.4.2 Likelihood based estimation

For a likelihood analysis the time-varying coefficient model proposed here is given by equation (5.3), and the parameters are estimated using the methods of Marx and Eilers (1998). They estimate (θ, α) by maximising the penalised log likelihood given by

$$\mathcal{L}(\theta, \alpha, \lambda | y) = \sum_{i=1}^{n} y_i \log(\mu_i) - \mu_i - \log(y_i!) - \lambda \theta^{\mathrm{T}} D \theta.$$

For notational clarity the time-varying coefficient model from equation (5.3) can be re-expressed as

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= \mathbf{M}_i^{\mathrm{T}} \delta, \end{aligned}$$

where $\delta = (\theta, \alpha)_{p+r \times 1}$ and $M = (\mathbf{M}_1^{\mathrm{T}}, \ldots, \mathbf{M}_n^{\mathrm{T}}) = (WB, Z)_{n \times p+r}$. Here $W = \text{diag}(w_1, \ldots, w_n)_{n \times n}$, B is the $n \times p$ basis matrix for the natural cubic spline and Z is the matrix of covariates. Consequently the penalised log likelihood can be re-expressed as

$$\mathcal{L}(\delta, \lambda | y) = \sum_{i=1}^{n} y_i \log(\mu_i) - \mu_i - \log(y_i!) - \lambda \delta^{\mathrm{T}} P \delta,$$

where $P = \begin{pmatrix} D_{p \times p} & \mathbf{0}_{p \times r} \\ \mathbf{0}_{r \times p} & \mathbf{0}_{r \times r} \end{pmatrix}$ and $\mathbf{0}$ denotes a matrix of zeros. The penalised likelihood can be maximised for specific values of λ using an iteratively re-weighted least squares algorithm, similar to that used in generalised linear models. The $(j+1)$th estimate of δ is given by

$$\delta^{(j+1)} = (M^{\mathrm{T}} K(\delta^{(j)}) M + \lambda P)^{-1} M^{\mathrm{T}} K(\delta^{(j)}) \tilde{\mathbf{y}}(\delta^{(j)}),$$

where $K(\delta^{(j)}) = \text{diag}(\mu_1(\delta^{(j)}), \ldots, \mu_n(\delta^{(j)}))$, $\tilde{\mathbf{y}}(\delta^{(j)}) = M\delta^{(j)} + K(\delta^{(j)})^{-1}(\mathbf{y} - \mu(\delta^{(j)}))$ and $\mu(\delta^{(j)}) = (\mu_1(\delta^{(j)}), \ldots, \mu_n(\delta^{(j)}))$. The asymptotic variance matrix is given by the sandwich estimator

$$\widehat{\mathrm{Cov}}(\hat{\delta}) = \hat{\phi}(M^{\mathrm{T}}K(\hat{\delta})M + \lambda P)^{-1}M^{\mathrm{T}}K(\hat{\delta})M(M^{\mathrm{T}}K(\hat{\delta})M + \lambda P)^{-1},$$

and the over-dispersion parameter is estimated as

$$\hat{\phi} = \frac{1}{n - \mathrm{tr}(H)} \sum_{i=1}^{n} \frac{(y_i - \mu_i(\hat{\delta}))^2}{\mu_i(\hat{\delta})}.$$

Here tr() denotes the trace of a matrix, while H denotes the 'Hat' matrix which is given by $H = M(M^{\mathrm{T}}K(\hat{\delta})M + \lambda P)^{-1}M^{\mathrm{T}}K(\hat{\delta})$. The trace of the 'Hat' matrix is the approximate number of parameters in the model, meaning that $\hat{\gamma}$ is approximately represented by $\mathrm{tr}(H) - r$ parameters. The parameter estimates and corresponding covariance matrix are dependent on the smoothing parameter λ, which can be estimated by optimising a data driven criterion such as generalised cross-validation (GCV) or Akaike's information criterion (AIC).

5.5 Simulation study

The simulation study presented in this section examines whether the time-varying coefficient model proposed in section 5.3 can adequately estimate different shaped temporally varying pollution-mortality relationships. Specifically four sets of mortality data are simulated, whose time-varying relationships with air pollution are (a) constant, (b) seasonal with a period of a year, (c) a quadratic trend, (d) a slowly evolving trend, which represent the type of relationships likely to be seen in real data. The Bayesian and likelihood implementations of this model are applied to each set of mortality data twice, firstly with the set of covariates used to simulate the data and secondly with a set chosen by model building criteria. The exact set of covariates are used to assess whether the proposed model produces accurate estimates in favourable conditions, while the latter mirrors the realistic situation where the set of confounding factors is unknown. The remainder of this section describes the simulation of the mortality data and presents the results.

5.5.1 Description of the data

In order to produce a realistic data set, each mortality series comprises three years of observations (1095 days) generated from a Poisson generalised linear model using PM_{10} and temperature data from one of the cities analysed in section 5.6 (Detroit). The vector of Poisson mean values depends on a one day lag of PM_{10} and a vector of covariates, the latter of which include an intercept term, pairs of sine and cosine terms with periods of a whole, half and a quarter of a year, and a natural cubic spline of the same days temperature with 3 degrees of freedom. The intercept term is fixed at 3.9 so that the expected number of daily deaths is between 30 and 70, which is similar to the real data analysed in section

5.6. The only differences between the four data sets are their time-varying relationships with PM_{10}, which are generated as follows:

(a) $\gamma_i = \theta_1$;

(b) $\gamma_i = \theta_1 + \theta_2 \sin(\omega_1 i) + \theta_3 \cos(\omega_1 i)$;

(c) $\gamma_i = \theta_1 + \theta_2 i + \theta_3 i^2$;

(d) $\gamma_i = \sum_{k=1}^{6} \theta_k B_k(i)$.

The last of these represents a slowly evolving long-term trend, which is modelled by a natural cubic spline with six degrees of freedom. The parameters that determine the pollution-mortality association are fixed so that the relative risks are around 1% (1.01), which is similar to those found in real data.

5.5.2 Results

The Bayesian estimates are based on 20,000 samples from the joint posterior distribution, which are generated using the MCMC algorithm described in section 5.4.1 (a single Markov chain is run for 130,000 iterations, the first 30,000 of which are discarded as burn-in while the remaining 100,000 are thinned by five to reduce their autocorrelation). The results are sensitive to the choice of inverse-gamma(ϵ, ϵ) prior for σ^2, as the pollution-mortality relationship becomes more variable as ϵ increases. This is because σ^2 is likely to be very small, and as ϵ increases it gives minimal or no prior weight to such values. As a result I adopt the approach suggested by Gelman (2006), and assign σ^2 a flat prior on the scale of standard deviation because it does not force σ^2 away from zero. In the likelihood analysis $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ are estimated by the penalised maximisation algorithm described in section 5.4.2, while λ is estimated by generalised cross validation (although Akaike's information criterion gives almost identical results). The results obtained with the exact covariates and those chosen by model building criteria are described below.

(i) - Exact set of covariates

The models are applied to the four simulated data sets using the same covariates that generated the mortality data. The results are shown in Figure 5-1, where the true relationships are represented by solid lines while the Bayesian and likelihood estimates are denoted by dotted and dashed lines respectively. The four panels relate to the different shaped time-varying relationships: (a) constant, (b) seasonal, (c) quadratic trend, (d) slowly evolving trend, and all the results are shown on the relative risk scale for an increase of $10\mu g/m^3$ in PM_{10}. The Bayesian and likelihood models estimate the temporal variation in the pollution-mortality relationships accurately, showing the correct overall shape in all

four cases. However both are less accurate if the relationship exhibits greater curvature, which can be seen by comparing panels (b) and (c). The Bayesian and likelihood estimates are very similar, and when they are different neither is preferable in general. In panel (b) the likelihood estimate appears to capture the seasonal behaviour better, while in panel (c) the Bayesian estimate is closer to the true quadratic trend.

(ii) - Covariates chosen by model building criteria

The four data sets are also analysed with a set of covariates chosen by model building criteria, which mirrors the standard situation where the confounding factors are unknown. These factors are modelled with an intercept term, a natural cubic spline of calendar time with fifteen degrees of freedom (five per year) and a spline of temperature with three degrees of freedom, which were chosen by DIC and residual based methods. The results are shown in Figure 5-2, which has an identical layout to the previous figure. The results are similar to those obtained using the exact set of covariates, as the Bayesian and likelihood estimates are similar and exhibit the correct underlying shape for all four data sets. A comparison of Figures 5-1 and 5-2 reveals that if the temporal variation in γ is small, both Bayesian and likelihood estimates are less accurate if the covariates are chosen by model building criteria. This can be seen by comparing panels (a) and (d) in the two figures, which shows that if the confounding factors are unknown the constant relationship in panel (a) is estimated as a slowly increasing long-term trend similar to that seen in panel (d). However the range of temporal variation estimated by both Bayesian and likelihood approaches is approximately 0.03 in panel (d), which is three times the corresponding value for panel (a). This suggests that an estimated pollution-mortality relationship that has a range of temporal variation less than 0.01 is indistinguishable from constant, while one that exhibits greater variation is unlikely to be constant.

5.6 Case study

This case study investigates the temporal variation in pollution-mortality relationships across four U.S. cities, using the time-varying coefficient model proposed in section 5.3. A number of simpler alternatives are also applied to these data, which allows an assessment of whether the additional complexity introduced by the proposed model is required. The first subsection describes the data that are used in this study, the second outlines the choice of statistical models, while the third presents the results.

5.6.1 Description of the data

The data used in this case study comprise daily measurements of mortality, pollution and meteorological covariates from four U.S. cities, that were first analysed by Samet et al. (2000) as part of the National Morbidity, Mortality and Air Pollution Study (NMMAPS).

Figure 5-1: Time-varying relationships between PM_{10} and mortality using the exact set of covariates that generated the health data. The true relationships are represented by solid lines while the Bayesian and likelihood estimates are shown as dotted and dashed lines respectively. The results are shown on the relative risk scale and the four panels depict relationships that are: (a) constant, (b) seasonal, (c) quadratic, (d) slowly evolving trend.

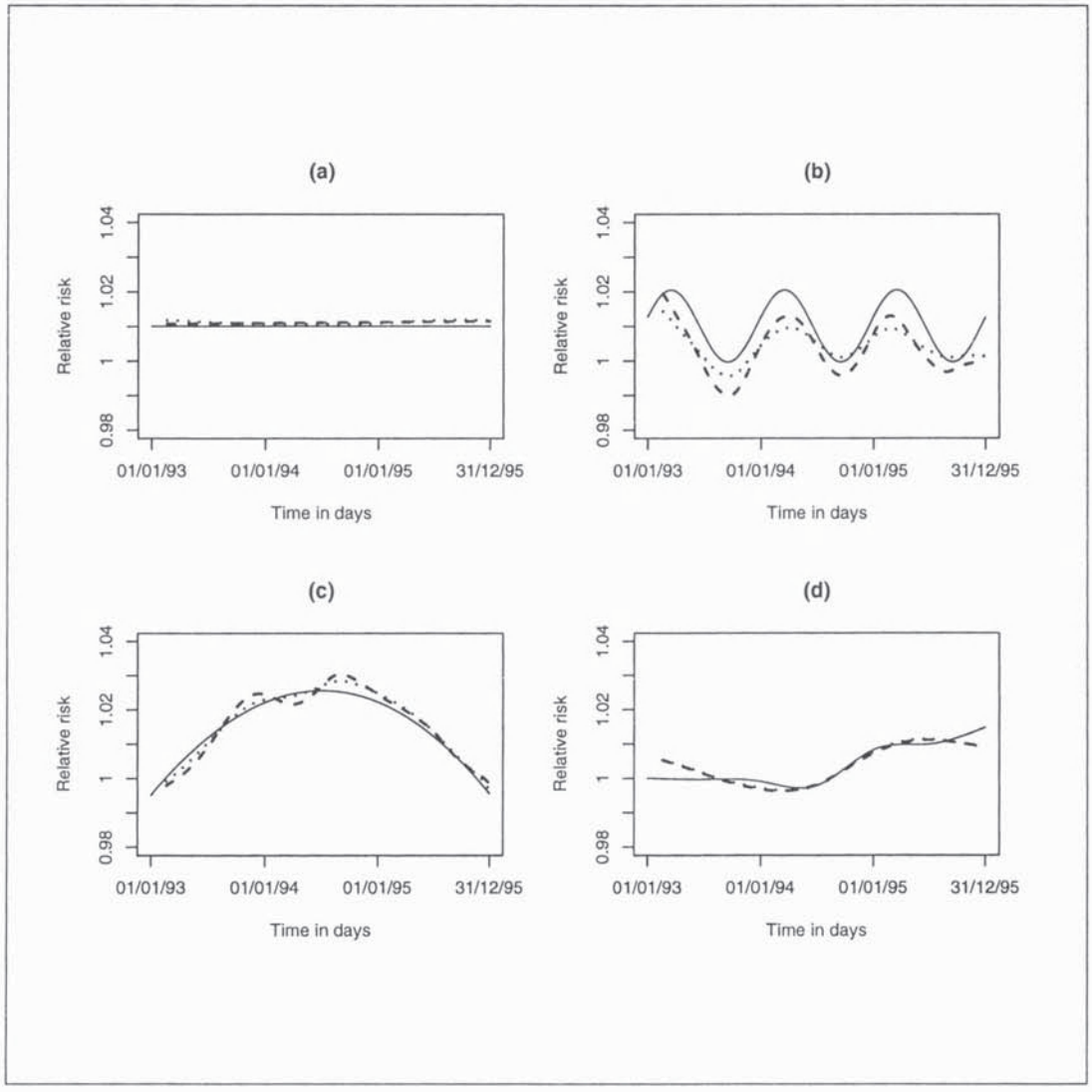
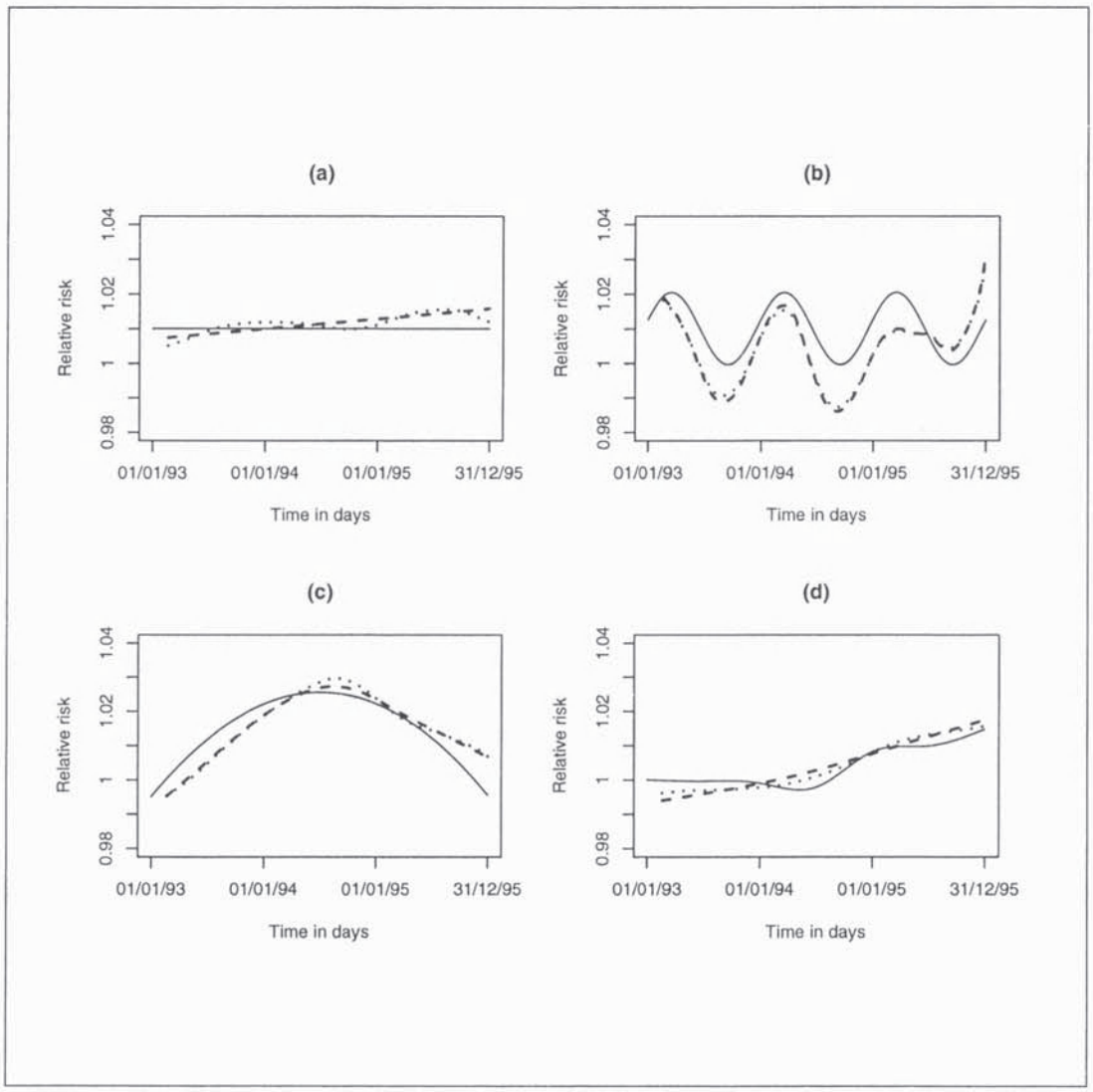


Figure 5-2: Time-varying relationships between PM_{10} and mortality using covariates chosen by model building criteria. The true relationships are represented by solid lines while the Bayesian and likelihood estimates are shown as dotted and dashed lines respectively. The results are shown on the relative risk scale and the four panels depict relationships that are: (a) constant, (b) seasonal, (c) quadratic, (d) slowly evolving trend.



These data were obtained from the R package 'NMMAPSdata' (Peng and Welty (2004)), which list mortality, pollution and meteorological data from 100 U.S. cities for fourteen years (1987 - 2000). This study analyses data from Cleveland, Detroit, Minneapolis and Pittsburgh from the 1st January 1993 until the 31st December 1997, which are chosen because they include the fewest missing values. The health data comprise daily counts of total non-accidental mortality for all age groups and are shown in Figure 5-4. Alternative mortality classifications (such as respiratory mortality) are also available, but they exhibit less than 10 events per day making them inappropriate for this type of study. The air pollution data comprise particulate matter levels measured as PM_{10} which are shown in Figure 5-3. These levels are calculated by averaging the measurements across a number of monitors in each city, a technique which is also applied to the Greater London data in chapter four. The meteorological covariates include mean temperature and dew point temperature (not shown), which exhibit seasonal patterns similar to those seen in the Greater London data.

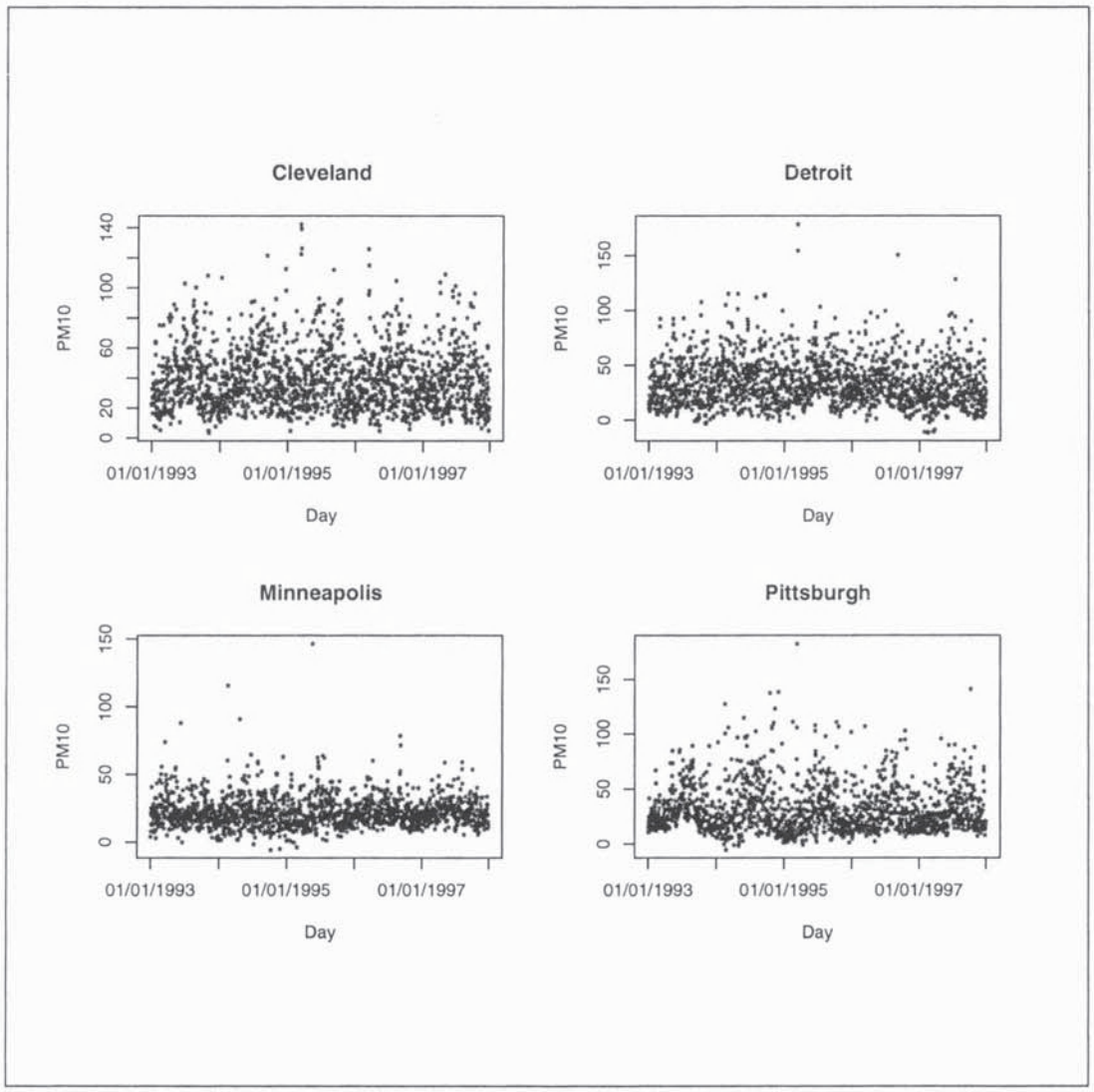
5.6.2 Description of the statistical models

To investigate the time-varying relationships between PM_{10} and total non-accidental mortality across the four cities, a series of six models are applied to each data set. These include the Bayesian and likelihood implementations of the time-varying coefficient model proposed in section 5.3, as well as a selection of alternatives adopted in existing studies. The general form of all six models is given by

$$\begin{aligned}
y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, 1826, \\
\ln(\mu_i) &= \text{PM}_{10_{i-1}}\gamma_i + \alpha_1 + S(i|35, \alpha_2) + S(\text{temperature}_0|6, \alpha_3) + S(\text{temperature}_{1-3}|6, \alpha_4) \\
&\quad + S(\text{dew point}_0|3, \alpha_5) + S(\text{dew point}_{1-3}|6, \alpha_6) + \text{DOW}_i\alpha_7, \\
\gamma_i &= g(i|\theta, \sigma^2), \\
(\theta, \sigma^2) &\sim f(\theta, \sigma^2), \\
\alpha &\sim \text{N}(\mu_\alpha, \Sigma_\alpha),
\end{aligned} \tag{5.7}$$

although the likelihood implementation of the time-varying coefficient model does not include the priors on the last two lines. In the model above $S(var|df)$ denotes a natural cubic spline of the variable 'var' with 'df' degrees of freedom. The covariates include an intercept term, six indicator variables for day of the week (denoted by 'DOW') and natural cubic splines of calendar time, daily mean temperature and dew point temperature. The meteorological covariates are included at a lag of 0 and as an average of the values from the previous three days, which are denoted by subscripts $_0$ and $_{1-3}$ respectively. The parameters that represent the covariates are collectively denoted by $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7)_{60\times1}$, which is assigned a non-informative multivariate

Figure 5-3: Particulate matter (PM_{10}) data from Cleveland, Detroit, Minneapolis and Pittsburgh.



Gaussian prior. The time-varying relationships between PM_{10} and mortality are determined by the function $g(i|\theta,\sigma^2)$ and the prior $f(\theta,\sigma^2)$, which are outlined below for each of the six models.

Model 1 Bayesian time-varying coefficient model - $g(i|\theta,\sigma^2)$ is represented as a penalised natural cubic spline:

$$\begin{aligned} \gamma_i &= \mathbf{B}_i^{\mathrm{T}}\theta, \\ \theta_l &\sim \mathrm{N}(2\theta_{l-1}-\theta_{l-2},\sigma^2) \quad \text{for } l=3,\ldots,p, \\ f(\sigma) &\propto 1. \end{aligned}$$

Estimation is achieved using the methods outlined in section 5.4.1.

Model 2 Likelihood time-varying coefficient model - $g(i|\theta,\sigma^2)$ is represented as a penalised natural cubic spline, meaning that $\gamma_i = \mathbf{B}_i^{\mathrm{T}}\theta$. The parameters that make up the spline are assigned a second order difference penalty, while the amount of smoothing is determined by λ. Estimation is achieved using the methods outlined in section 5.4.2.

Model 3 Stochastic model - $g(i|\theta,\sigma^2)$ is represented as a first order random walk:

$$\begin{aligned} \gamma_i &\sim \mathrm{N}(\gamma_{i-1},\sigma^2), \\ \sigma^2 &\sim \mathrm{N}(0,g_1)_{I[\sigma^2>0]}. \end{aligned}$$

This is the model first adopted by Chiogna and Gaetan (2002), which was subsequently presented in chapter four. It is implemented within a Bayesian setting using the algorithm described in that chapter.

Model 4 Trend model - where $g(i|\theta,\sigma^2)$ is represented as a cubic polynomial

$$\begin{aligned} \gamma_i &= \theta_1+\theta_2 i+\theta_3 i^2+\theta_4 i^3, \\ \boldsymbol{\theta} &\sim \mathrm{N}(\mu_\theta,\Sigma_\theta). \end{aligned}$$

This is a simpler alternative for modelling relationships that exhibit long-term trends.

Model 5 Seasonal model - $g(i|\theta,\sigma^2)$ is represented as a cyclical function:

$$\begin{aligned} \gamma_i &= \theta_1+\theta_2\sin(2\pi i/365)+\theta_3\cos(2\pi i/365), \\ \boldsymbol{\theta} &\sim \mathrm{N}(\mu_\theta,\Sigma_\theta). \end{aligned}$$

This is the model used by Peng et al. (2005) to investigate whether the pollution-mortality relationship exhibits seasonal variation.

Model 6 Constant model - $g(i|\theta, \sigma^2)$ is fixed at a single value:

$$\begin{aligned} \gamma_i &= \theta_1, \\ \theta_1 &\sim \mathrm{N}(0, 10). \end{aligned}$$

This is the standard air pollution and mortality relationship, which is used to emphasize the temporal variation (or lack of it) estimated by the other five models.

Models four to six simplify to generalised linear models, meaning that Bayesian estimation can be implemented using the MCMC algorithm outlined in chapter 2.3.

Model building strategy

Model building is implemented within a Bayesian framework using the deviance information criteria. Each set of mortality data exhibit long-term trends, seasonal variation, over-dispersion and temporal correlation, which are typically modelled with smooth functions of calendar time and meteorological covariates such as temperature. In this study we retain this standard approach to confounder control, even though chapter four shows that a second order random walk is preferable to a smooth function of calendar time. In common with that chapter the smooth functions are represented as natural cubic splines, because they are straightforward to implement within Bayesian and likelihood settings.

The model building process began by fitting natural cubic splines of calendar time to each data set, with degrees of freedom ranging between 5 and 60. Thirty five degrees of freedom are chosen for all four cities, because it removes the trend and temporal correlation present in the mortality series and has the lowest DIC in three of the four cases (in Detroit it was nearly optimal). This function also removes part of the seasonal variation present in the mortality data, while the remainder is typically modelled by natural cubic splines of meteorological covariates. For these cities both temperature and dew point temperature data are available, and have been used for confounder control in a number of existing studies (see for example Dominici et al. (2000)). The relationships between both covariates and mortality were investigated at a number of lags, and values on the same day and a moving average over the previous three days were found to be significant (posterior credible intervals do not include zero) for all four cities. The relationship between each of these covariates and mortality was investigated using DIC, and a smooth function with a low degrees of freedom is adopted in all cases. The smooth functions of temperature and dew point temperature are modelled with six and three degrees of freedom respectively,

because these covariates exhibit U-shaped relationships with mortality. A set of indicator variables for day of the week are also included as covariates, because they show a significant relationship with mortality and have been used in previous studies. The vector of covariate parameters α is assigned a non-informative multivariate Gaussian prior, which has a diagonal variance matrix and a mean that is based on data from earlier years (1990-1992).

After removing the influence of the confounding factors PM_{10} was added to the models. A number of lags and moving averages were fitted to each data set, and a lag of one day is used in all cases. This is chosen because it has the lowest DIC for three of the four cities, exhibits a significant posterior estimate in all cases and has been used in previous analyses of these data (Dominici et al. (2000)). For simplicity each lag was investigated assuming its relationship with mortality was constant (Model 6). For the time-varying coefficient models (Models 1 and 2) a sensitivity analysis was conducted to determine the number of basis functions required to adequately represent the PM_{10}-mortality relationships. Between 60 and 100 basis function were examined for each data set, and as the results showed little change 60 functions (12 per year) are used in the final models.

For the Bayesian time-varying coefficient model (Model 1) a sensitivity analysis was conducted to determine whether the choice of prior for σ^2 affected the results. A range of inverse-gamma(ϵ, ϵ) priors were applied to each data set, and the results exhibited substantial heterogeneity. This variation is caused by the small values of σ^2, which are estimated to be between 10^{-7} and 10^{-6} if an improper non-informative prior is used. However the 'non-informative' priors described above have little or no mass at such small values, meaning that σ^2 increases as ϵ gets bigger which leads to a less smooth estimate of γ. As the question of interest is whether such temporal variation exists, adopting an inverse-gamma(ϵ, ϵ) prior which forces σ^2 away from zero is not appropriate, because it will bias against a constant association. As a result I adopt the approach suggested by Gelman (2006), and use a flat prior on the scale of standard deviation (effectively an improper inverse-gamma(-0.5,0) distribution). This prior has significant mass close to zero and gives results that are similar to those obtained as $\epsilon \to 0$. A similar sensitivity analysis was carried out for the random walk model (Model 3), and the time-varying relationship was highly dependent on the choice of prior for σ^2. If a non-informative prior is used the estimated relationship is contaminated with unwanted noise, meaning that an underlying trend cannot be seen. As a result I adopt the approach described in chapter four, and assign σ^2 an informative zero mean Gaussian prior that is truncated to be positive. The variance of this prior is chosen to ensure the relationship is smooth, which results in $g_1 = 10^{-12}$. The remaining three models (Models 4 to 6) simplify to generalised linear models, and non-informative Gaussian priors are specified for θ.

5.6.3 Results

The models contain a large number of covariates, so to aid convergence they are standardised to have a mean of zero and a standard deviation of one before inclusion in the model (and are subsequently back-transformed when obtaining results from the posterior distribution). The Bayesian analyses are based on 50,000 samples from the joint posterior distribution, which are generated by two Markov chains. Each chain is 'burnt in' for 40,000 iterations, at which point convergence was assessed to have been reached using the methods of Gelman et al. (2003). A further 125,000 samples are then generated from each chain, which are thinned by 5 to reduce the correlation. For the likelihood implementation of the time-varying coefficient model (Model 2) the smoothing parameter λ is estimated by GCV, although AIC gives identical results. Figure 5-4 shows the observed and estimated daily mortality counts for each city, with the latter coming from the Bayesian implementation of the time-varying coefficient model (Model 1). For each data set the seasonal trend in mortality is well estimated while the residuals exhibit little structure or correlation (not shown), indicating that the chosen covariates adequately model the unknown risk factors that afflict these data. The remainder of this section investigates the time-varying relationships between PM_{10} and mortality in each city, examining the performance of the time-varying coefficient model proposed here and the simpler alternatives adopted in the related literature. For brevity in the following discussion each model is referred to by the function that estimates its time-varying relationship, resulting in the following model names: Bayesian spline (Model 1), likelihood spline (Model 2), random walk (Model 3), cubic (Model 4), seasonal (Model 5), and constant (Model 6).

(i) - Bayesian and likelihood spline models

The time-varying relationships between PM_{10} and mortality for all four cities are shown in Figures 5-5 (Bayesian spline) and 5-6 (likelihood spline), which are presented on the relative risk scale for an increase of $10\mu g/m^3$ in PM_{10}. The solid lines depict the time-varying estimates from these models, the grey shading represents pointwise 95% credible or confidence intervals, while the dashed lines represent the constant relationships (Model 6) which are shown for comparison. All the time-varying estimates exhibit long-term trends, which can be increasing (Detroit) or decreasing (Pittsburgh) over the five year period. Their overall size is approximately 1.005, which is similar to the results found in previous analyses of these data (see for example Samet et al. (2000)). The defining characteristics of these relationships are summarised in Table 5.1, which gives their minimum and maximum values as well as their ranges. The ranges are approximately 1% in Cleveland and Detroit, between 1% and 2% in Minneapolis, and 2% in Pittsburgh, indicating that with the exception of Minneapolis the Bayesian and likelihood estimates exhibit similar amounts of temporal variation. All eight estimates have the appearance of low order polynomials, a fact that is re-enforced by the estimated smoothing parameters given in Table 5.2. The

Figure 5-4: Observed (stars) and estimated (solid line) mortality counts from Cleveland, Detroit, Minneapolis and Pittsburgh. The estimated values are from the Bayesian spline model (Model 1).

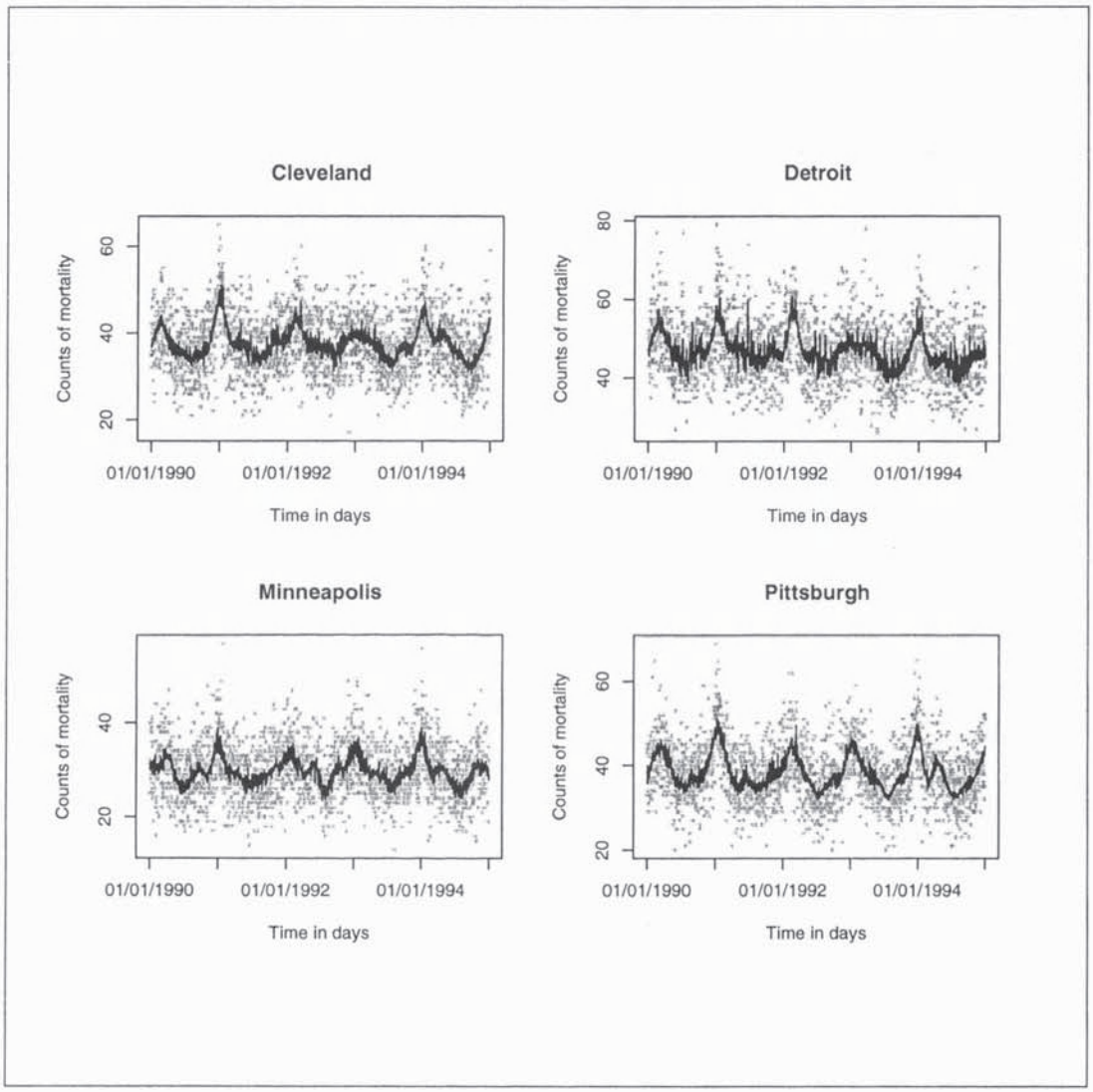


Table 5.1: Summary of the time-varying relative risks. The minimum and maximum values are in brackets with the range given below.

Model	City			
	Cleveland	**Detroit**	**Minneapolis**	**Pittsburgh**
1	(0.999 - 1.009) 0.010	(1.000 - 1.011) 0.011	(0.998 - 1.015) 0.026	(0.993 - 1.011) 0.018
2	(0.998 - 1.007) 0.009	(1.000 - 1.009) 0.009	(1.000 - 1.009) 0.009	(0.995 - 1.012) 0.017
3	(1.000 - 1.008) 0.008	(1.002 - 1.007) 0.005	(1.001 - 1.010) 0.009	(0.997 - 1.011) 0.014
4	(0.996 - 1.010) 0.014	(1.002 - 1.008) 0.006	(0.966 - 1.030) 0.064	(0.989 - 1.020) 0.031
5	(0.997 - 1.012) 0.015	(0.999 - 1.011) 0.012	(1.003 - 1.007) 0.004	(1.001 - 1.008) 0.007
6	1.005	1.005	1.005	1.005

smoothness of the likelihood estimates are summarised by their approximate degrees of freedom, which are between two and three, the same amount used to model a low order polynomial. This indicates that the smoothness of the time-varying relationships is close to maximal, because natural cubic splines have the appearance of low order polynomials in this case. The Bayesian relationships also exhibit close to maximal smoothness, as the estimated variances are all close to zero. For each city the Bayesian and likelihood estimates are very similar, although the former tend to exhibit greater curvature than their likelihood counterparts. The biggest difference between the two estimates is observed in Minneapolis, where the ranges are 1.7% (Bayesian) and 0.9% (likelihood) respectively. These differences are most likely a result of the techniques used to estimate the respective smoothing parameters, a point which is taken up in the discussion. The Bayesian credible intervals are wider than the corresponding likelihood confidence intervals, which is also a by-product of the way σ^2 and λ are estimated. This is discussed further in section 5.7, where it is argued that the likelihood confidence intervals are likely to be too narrow. The width of both credible and confidence intervals vary between the four cities, with Minneapolis exhibiting intervals that are unusually large.

(ii) - Comparison with previously adopted approaches

The time-varying relationships estimated by models three to six are shown in Figures 5-7 and 5-8, the first of which presents the results for the random walk model (Model 3) while the latter displays the three simple parametric forms (Models 4 to 6). The random walk estimates are very similar to those obtained from the Bayesian and likelihood spline models, particularly the Bayesian implementation. The 95% credible intervals from the

Figure 5-5: Estimated time-varying relationships between PM_{10} and mortality from the Bayesian spline model (Model 1). The grey shading represents 95% credible intervals while a dashed line depicts a constant relationship (Model 6).

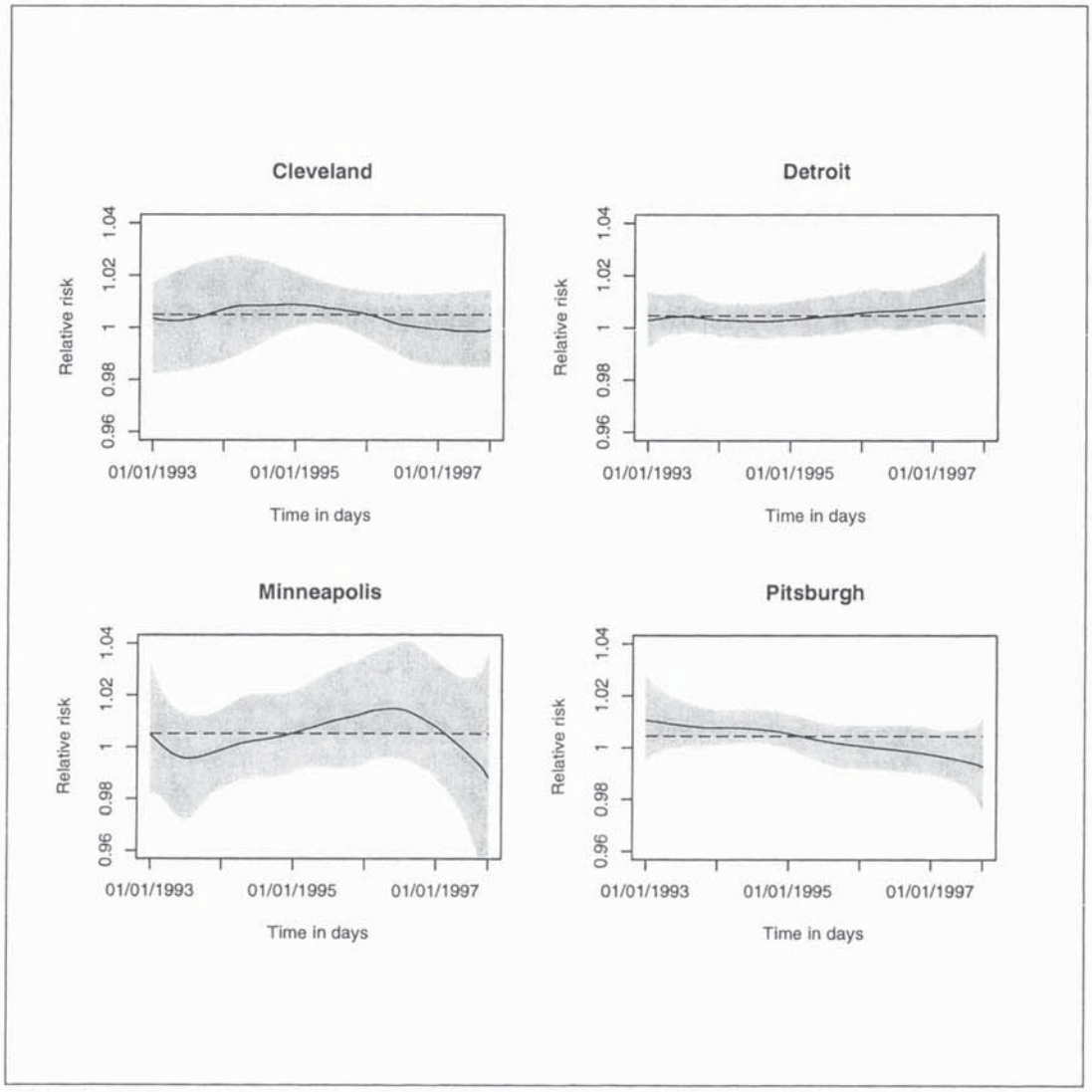
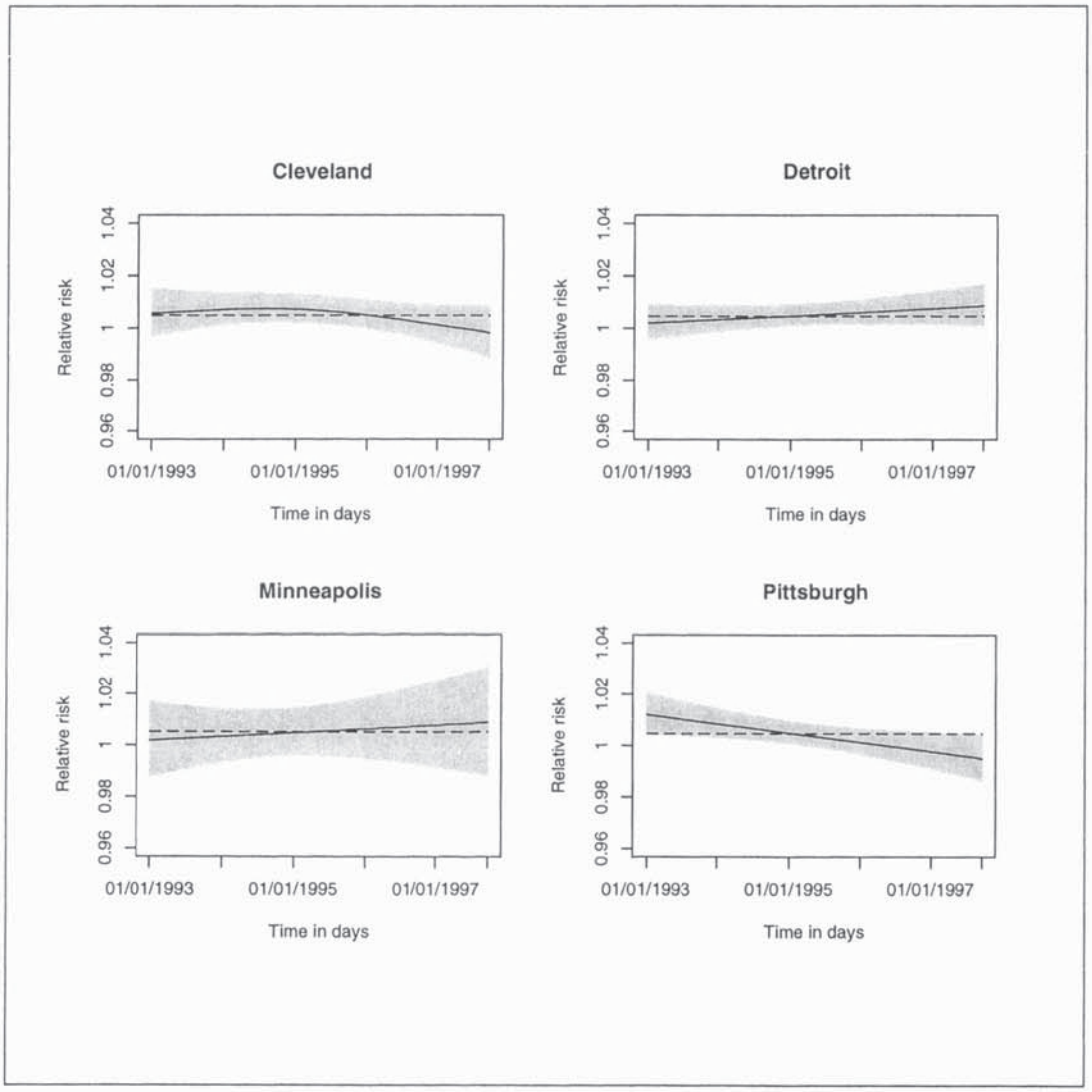


Table 5.2: Summary of the smoothing parameters.

Model and parameter	City			
	Cleveland	Detroit	Minneapolis	Pittsburgh
1 - σ^2	1.1×10^{-6}	5.3×10^{-7}	1.7×10^{-6}	6.1×10^{-7}
2 - degrees of freedom	2.6	2.0	2.0	2.0
3 - σ^2	4.8×10^{-7}	3.7×10^{-7}	4.0×10^{-7}	4.6×10^{-7}

Figure 5-6: Estimated time-varying relationships between PM_{10} and mortality from the likelihood spline model (Model 2). The grey shading represents 95% confidence intervals while a dashed line depicts a constant relationship (Model 6).



random walk and Bayesian spline models are also similar, which is not surprising given the similarities in the way they are calculated. The cubic model also produces similar estimates to models one to three, indicating that if the relationship exhibits a long-term trend this simple approach is adequate. However, except for Detroit the cubic estimates exhibit greater curvature than those from the more complex approaches (more than double the range), which is at odds with its comparatively simple parametric form. The seasonal estimates (Model 5) are completely at odds with those obtained from the flexible models, suggesting that they are entirely spurious and caused by the models restrictive nature.

5.7 Discussion

This chapter examines whether the relationship between air pollution and mortality exhibits any temporal variation, and proposes a time-varying coefficient model for this purpose. The efficacy of this model is assessed via a simulation study, where it is applied to data sets with different shaped time-varying relationships. It is then used to investigate the pollution-mortality associations in four U.S. cities, and the results are compared against alternative models that have been proposed in the related literature. In the four cities studied the relationship between PM_{10} and total non-accidental mortality is approximately 1.005 in size (given by the constant model), which is similar to results presented in previous analyses of these data (see for example Samet et al. (2000)). The time-varying relationships exhibit both increasing (Detroit) and decreasing (Pittsburgh) long-term trends, but the range of temporal variation is relatively small, being approximately 1% in Cleveland and Detroit, between 1% and 2% in Minneapolis (depending on the chosen model), and 2% Pittsburgh. However this variation only relates to a five year interval (1993 - 1997), and it is of interest to examine whether these trends continue over a longer period of time.

The time-varying coefficient model proposed in this chapter represents the temporal evolution in the pollution-mortality relationship as a penalised natural cubic spline, because it produces estimates that are smooth with no fixed shape. The simulation study shows that the Bayesian and likelihood implementations of this model estimate a variety of different shaped time-varying relationships closely, including constant, seasonal and long-term trends. The models are most accurate if the true relationship shows little curvature and the confounding factors afflicting the mortality data are known, although in other situations the models produce estimates that retain their correct overall shape. However, if the confounding factors are unknown a constant association may be indistinguishable from a slowly evolving trend, whose maximum and minimum relative risks differ by less than 1%. The time-varying relationships estimated by the Bayesian and likelihood penalised spline models are very similar, and the simulation study indicates that neither are more accurate in general. However the former typically exhibit greater curvature, which is probably due to the methods used to estimate the respective smoothing parameters (λ, σ^2). In the

Figure 5-7: Estimated time-varying relationships between PM_{10} and mortality from the random walk model (Model 3). The grey shading represents 95% credible intervals while a dashed line depicts a constant relationship (Model 6).

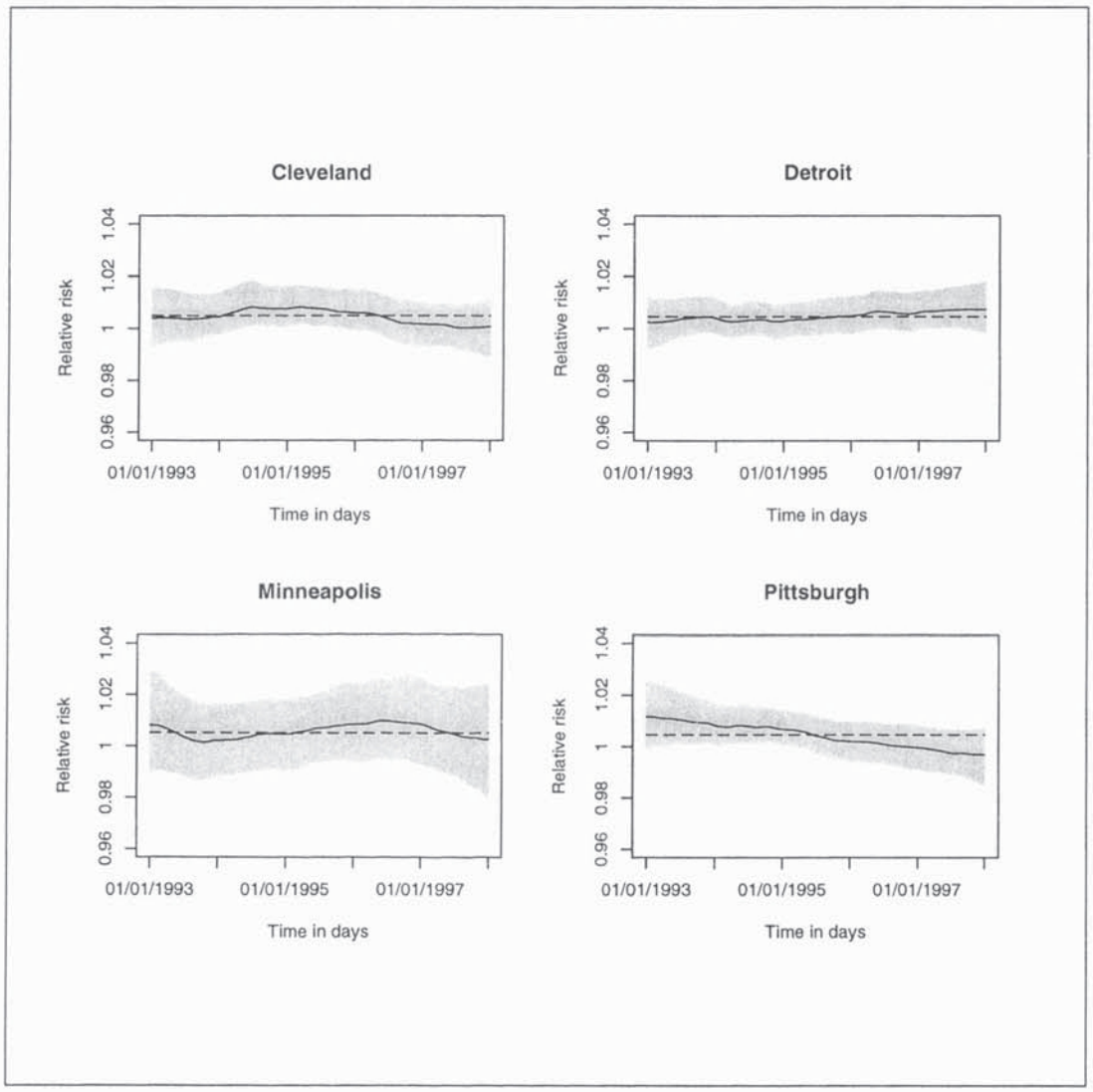
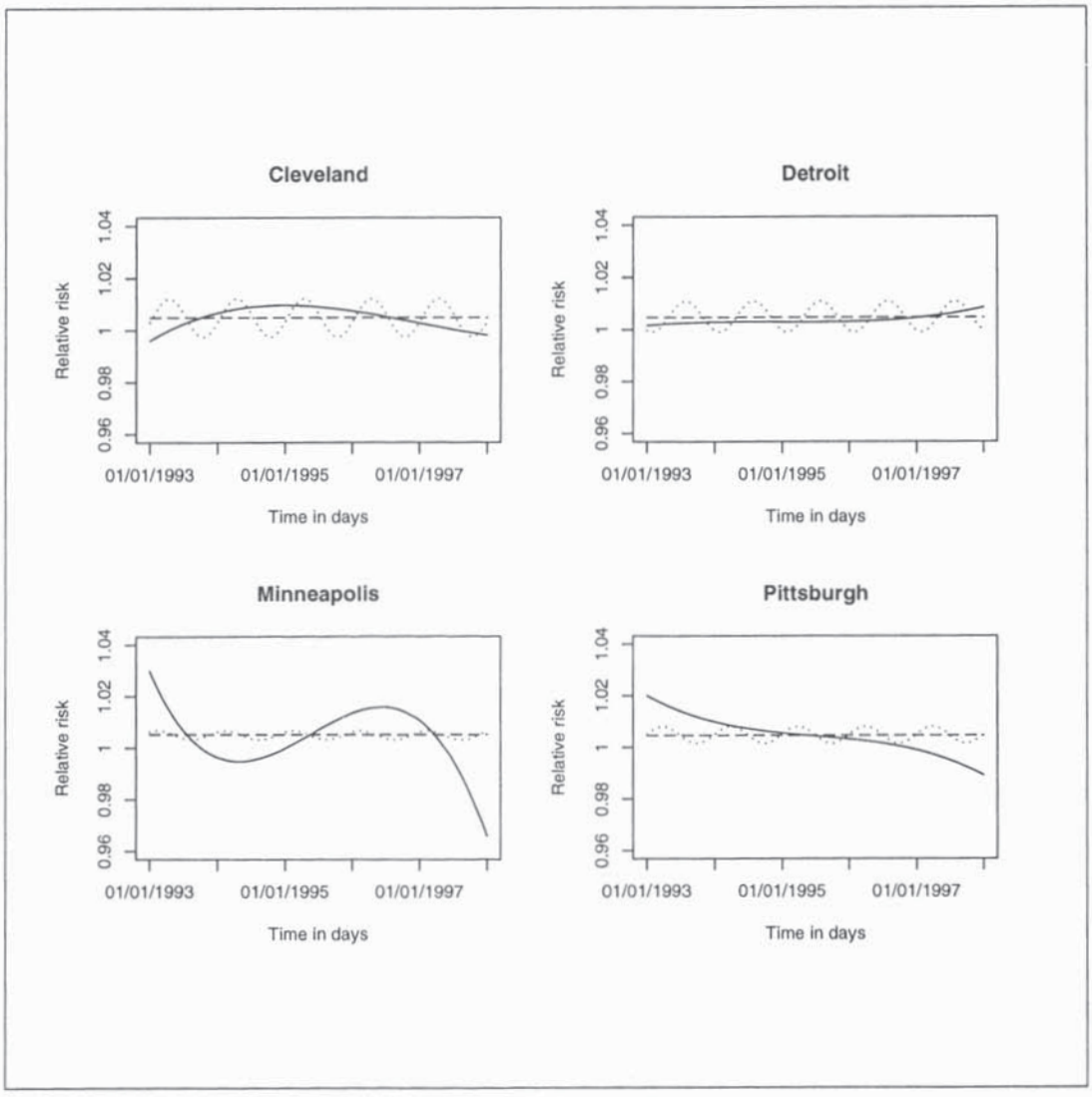


Figure 5-8: Estimated time-varying relationships between PM_{10} and mortality from the three parametric models. The cubic (Model 4), seasonal (Model 5) and constant (Model 6) relationships are represented by solid, dotted and dashed lines respectively.



likelihood implementation the time-varying relationship is estimated for numerous values of the smoothing parameter, which is then selected by optimising a model selection criterion. In contrast for the Bayesian analysis the smoothing parameter is estimated as part of the MCMC algorithm, meaning that the pollution-mortality relationship depends on its entire posterior distribution. This includes values that correspond to highly variable relationships, which causes the estimated association to be less smooth. The Bayesian credible intervals are generally wider than their likelihood counterparts, which is also due to the methods used to estimate the respective smoothing parameters. As described above the Bayesian intervals correctly incorporate the uncertainty in σ^2, where as the likelihood intervals are based on a fixed estimate of λ. Consequently the Bayesian intervals are more realistic estimates of the true variability, in comparison with their likelihood counterparts which are likely to be too narrow.

The efficacy of the proposed model is also assessed in the case study, where its time-varying estimates are compared against those from alternative models. These alternatives have been adopted in existing air pollution studies, and represent the pollution-mortality relationship with a random walk, a low order polynomial, a seasonal yearly cycle and a constant function. The best of these is the first order random walk, because unlike the simple parametric forms it does not restrict the shape of the estimated association. However its smoothness is highly dependent on $f(\sigma^2)$ (the prior for the smoothing parameter), with a constant association and noisy estimates both obtainable by particular prior specifications. The simple parametric forms are not appropriate representations of the pollution-mortality relationship, because they restrict the estimate to exhibiting a fixed parametric form. The relationships estimated in the case study exhibit long-term trends, meaning that the cubic model appears to perform well while the seasonal model produces spurious results. This suggests that the models of Moolgavkar et al. (1995) and Peng et al. (2005) are too restrictive for this problem, because they force the temporal evolution to follow rigid seasonal forms. In contrast the penalised spline model proposed here is more flexible, allowing the relationship to adopt a variety of temporally varying shapes that are beyond the scope of simple parametric models. The estimates from the cubic model exhibit greater curvature than those from the spline or random walk models, which initially seems at odds with its simpler parametric form. However the cubic model is not subject to any smoothing penalties or restrictions, meaning that its estimates are based on four degrees of freedom. In contrast the flexible models (spline and random walk) are smoothed to remove excess variation, which reduces their degrees of freedom and makes the estimates less variable. For example, Table 5.2 shows that the likelihood spline model estimates the time-varying relationship with less than three degrees of freedom, which compares with four used by the cubic trend.

This chapter has investigated the time-varying relationships between PM_{10} and mortality

in four U.S. cities, with particular focus on whether they exhibit any temporal variation. These relationships are consistently estimated as non-constant by the Bayesian and likelihood time-varying coefficient models, which the simulation study showed produce estimates that have the correct overall shape. However, the pointwise credible and confidence intervals are too wide to preclude the possibility that the true association is constant. In addition, the simulation study highlights that a constant association may be mis-estimated as a long-term trend that exhibits a small amount of temporal variation (up to 1% on the relative risk scale). Consequently there is insufficient evidence that the relationships in Cleveland, Detroit and Minneapolis change over time, while the 2% decrease in relative risk observed in Pittsburgh is more likely to be a true effect.

Chapter 6

Assessing the impact of mis-estimating air pollution exposure

6.1 Introduction

The relationship between air pollution and mortality is a central theme of the previous two chapters, with particular interest in whether it exhibits any temporal variation. In this chapter I retain the interest in the air pollution component, but focus on how pollution exposure is estimated. In particular I compare three alternative estimates of exposure, the first of which is adopted by the majority of existing air pollution studies. This 'standard' estimate is described in section 3.2, and is a daily average of the pollution data from all available monitoring sites. The second estimate is an intermediate approach using a measurement error model, while the third is based on a spatio-temporal pollution model. Each approach is compared and contrasted using both simulated and real data, with the analysis throughout being Bayesian, using Markov chain monte carlo simulation.

The remainder of this chapter is organised as follows. Section 6.2 reviews the background and motivation for this work, including a critique of the standard approach to estimating pollution exposure in ecological studies. The next two sections describe alternative models for estimating exposure, which incorporate the factors likely to bias the standard estimate. Section 6.3 describes a model that incorporates the possibility of measurement error in the ambient monitoring data, while section 6.4 presents a model that additionally accounts for spatial variation. Section 6.5 compares all three models via a simulation study, which investigates their relative performances for a variety of pollution surfaces. Section 6.6 extends this comparison to real data, by estimating the relationships between multiple pollutants and mortality in Greater London. Finally section 6.7 presents a concluding

discussion.

6.2 Background and motivation

6.2.1 Review of standard air pollution studies

The ecological studies described in this thesis are based on mortality, pollution and meteorological data, that relate to a fixed region $\mathcal{R}$ for n consecutive days. The daily mortality counts are drawn from the population living within $\mathcal{R}$, and are denoted here by $\mathbf{y} = (y_1, \ldots, y_n)_{n\times 1}$. These counts are regressed against air pollution levels and a vector of q covariates, the latter of which are denoted by $Z = (\mathbf{z}_1^{\mathrm{T}}, \ldots, \mathbf{z}_n^{\mathrm{T}})_{n\times q}$, where $\mathbf{z}_i^{\mathrm{T}} = (z_{i1}, \ldots, z_{iq})_{1\times q}$ are the realisations for day i. The pollution data are collected from k fixed site monitors located across the region of interest ($\mathcal{R}$), which measure ambient pollution levels continuously throughout the day. A daily average is typically calculated at each location, which for day i and location $\mathbf{s}_l$ is denoted by $w_i(\mathbf{s}_l)$. The set of pollution locations are collectively denoted by $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_k\}$ (where $\mathbf{s}_l = (a_l, b_l) \in \mathcal{R}$), and for day i the pollution levels are summarised by $\mathbf{w}_i(\mathcal{S}) = (w_i(\mathbf{s}_1), \ldots, w_i(\mathbf{s}_k))_{k\times 1}$. The entire set of pollution data are collected into an $n \times k$ matrix denoted by $W(\mathcal{S}) = (\mathbf{w}_1(\mathcal{S})^{\mathrm{T}}, \ldots, \mathbf{w}_n(\mathcal{S})^{\mathrm{T}})_{n\times k}$, which is likely to contain a small proportion (typically less than 10%) of missing values. In ecological studies of this type pollution exposure is typically estimated by $\mathbf{w} = (w_1, \ldots, w_n)$, where

$$w_i = \frac{1}{k}\sum_{l=1}^{k} w_i(\mathbf{s}_l) \quad \text{for } i = 1, \ldots, n, \tag{6.1}$$

the average value from the k monitors (missing values are typically ignored). The repeated use of this estimate in existing studies is due to its simplicity, being a simple average that requires no additional modelling. The relationship between $\mathbf{w}$ and mortality is typically estimated using Poisson regression models, such as those given in equations (3.1) and (3.2).

6.2.2 Limitations

The primary aim of the study described above is to estimate the relationship between ambient pollution levels and mortality or morbidity, which is represented by γ in equations (3.1) and (3.2). The accuracy of this estimate depends on the estimated exposure ($\mathbf{w}$) and the choice of covariates, the latter of which should ideally remove any structure, over-dispersion or underlying temporal correlation from the daily mortality series. In this chapter I investigate how the accuracy of $\mathbf{w}$ affects the estimation of γ, which requires the true pollution exposure to be known. Personal pollution exposures are typically unavailable (see chapter 3.8), so ambient levels measured by the fixed site monitors have to be used as a surrogate. However these pollution data are spatially misaligned in relation to

the mortality events, because the former are measured at k fixed locations while the latter are drawn from the population living throughout $\mathcal{R}$. Gelfand et al. (2001) call this the *change of support problem*, and argue that the desired exposure is the average ambient pollution level across $\mathcal{R}$. This is because it relates to the same geographical region as the mortality counts, and is denoted here by $\mathbf{x} = (x_1, \ldots, x_n)$. For day i this is defined as

$$x_i = \text{average}[x_i(\mathbf{s})] = \frac{1}{|\mathcal{R}|} \int_{\mathbf{s} \in \mathcal{R}} x_i(\mathbf{s}) d\mathbf{s} \quad \text{for } i = 1, \ldots, n, \tag{6.2}$$

where $|\mathcal{R}|$ denotes the size of $\mathcal{R}$, and $x_i(\mathbf{s})$ denotes the true ambient pollution concentration at spatial location $\mathbf{s}$ on day i. As described above the majority of authors estimate $\mathbf{x}$ with $\mathbf{w}$, and the amount of bias in this estimate will largely depend on two characteristics of the underlying pollution surface:

(a) - **Spatial variation** - If the underlying surface exhibits substantial spatial variation, the measurements at the monitor locations are unlikely to be a representative sample of the pollution levels throughout $\mathcal{R}$. This is because the monitor locations are likely to be small in number, unequally spaced throughout the data and located for specific reasons (for example at a well known pollution hot spot), meaning that averaging the values at these sites may not produce a good estimate of $\mathbf{x}$.

(b) - **Measurement error** - The ambient monitors are known to measure with error (Department of the Environment, Transport and the Regions (1998)), meaning that $\mathbf{w}$ may be a biased estimate of $\mathbf{x}$. In addition this error is likely to inflate the variance of γ, meaning that (3.2) and (6.1) will underestimate the uncertainty in this estimate.

If both these factors are negligible $\mathbf{w}$ is likely to be a reasonable estimate of $\mathbf{x}$, and the resulting estimate of γ should be close to the true association. However if either of these factors are non-negligible the impact on $\mathbf{w}$ and γ is unknown, and will require further investigation. A small number of authors (see for example Duddek et al. (1995) Carlin et al. (1999) and Zhu et al. (2003)) have replaced $\mathbf{w}$ with an alternative estimate, that incorporates spatial variation and measurement error by modelling the underlying pollution surface with a spatio-temporal model. However these authors focused on analysing data from a single region, and did not investigate whether the additional complexity of a spatio-temporal model had a noticeable impact on the results. In this chapter I address this question directly, and compare the performance of $\mathbf{w}$ against two alternative models that vary in complexity. The first of these incorporates the possibility of measurement error, but assumes the pollution surface is spatially flat. The second alternative is a spatio-temporal pollution model, which incorporates both spatial variation and measurement error and is similar to that used by Duddek et al. (1995), Carlin et al. (1999) and Zhu et al. (2003). This model would be expected to produce the most accurate estimates of

$\mathbf{x}$ from the three alternatives, because it directly models the underlying pollution surface. However this means it requires a large number of spatial observations to produce accurate estimates, and is the most computationally intensive to implement. Therefore the choice of pollution model requires a tradeoff between simplicity and accuracy, and an aim of this chapter is to examine in what situations each alternative is suitable. In particular the work presented in this chapter is motivated by three questions: (i) how accurately does the standard approach (w) estimate average pollution exposure and its relationship with mortality; (ii) what factors affect the accuracy of the standard approach, and in what situations should it be used; (iii) do the more complex alternatives investigated here produce significantly better estimates. These questions are addressed in sections 6.5 and 6.6, while the two alternative estimates of $\mathbf{x}$ are presented in sections 6.3 and 6.4. To ensure that the only difference between the three models is the method used to estimate $\mathbf{x}$, the same mortality model is used in all cases.

6.3 A measurement error model

A measurement error model is an intermediate step between the standard approach and a spatio-temporal pollution model, which may be a useful alternative if the pollution surface is spatially flat. The model assumes the true pollution surface is spatially constant (that is $x_i(\mathbf{s}) = x_i \ \forall \ \mathbf{s} \in \mathcal{R}$), which implies that $\mathbf{w}_i(\mathcal{S})$ is likely to be a vector of k independent estimates of x_i. This suggests that a classical measurement error model is appropriate, a brief review of which is presented in chapter 2.8. The model is based on a joint likelihood for $(\mathbf{y}, \mathbf{x}, W(\mathcal{S}))$, which can be decomposed into

$$f(\mathbf{y}, \mathbf{x}, W(\mathcal{S})|\boldsymbol{\Omega}) = f(\mathbf{y}|\mathbf{x}, \boldsymbol{\omega_1}) f(W(\mathcal{S})|\mathbf{x}, \boldsymbol{\omega_2}) f(\mathbf{x}|\boldsymbol{\omega_3}), \tag{6.3}$$

where $\boldsymbol{\Omega} = (\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \boldsymbol{\omega_3})$ is a vector of parameters. The measurement error is assumed to be non-differential for simplicity, meaning that $\mathbf{y}$ and $W(\mathcal{S})$ are conditionally independent given $\mathbf{x}$. The Bayesian measurement error model is given by

$$\begin{aligned}
y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\
\ln(\mu_i) &= x_i\gamma + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\alpha}, \\
\boldsymbol{\beta} = (\gamma, \boldsymbol{\alpha}) &\sim \text{N}(\boldsymbol{\mu}_\beta, \Sigma_\beta), \\
\ln(w_i(\mathbf{s}_l)) &\sim \text{N}(\ln(x_i), \sigma_\epsilon^2) \quad \text{for } l = 1, \ldots, k, \\
\ln(x_i) &\sim \text{N}(\mu_x, \sigma_x^2), \\
\sigma_\epsilon^2 &\sim \text{inverse-gamma}(e, f).
\end{aligned} \tag{6.4}$$

The first term in equation (6.3) is represented by a Poisson generalised linear model (specified by the first three lines of equation (6.4)), which is identical to the standard

model except that w_i is replaced by x_i. The second term in the joint likelihood is a classical measurement error model on the natural log scale, which is represented by the fourth line in equation (6.4). The natural log scale is used here as air pollution data are likely to arise from a log-normal distribution (Ott (1990)), because they are non-negative measurements that may have a skewed distribution. The last term in equation (6.3) acts as a prior for $\ln(\mathbf{x})$, which is specified as a non-informative Gaussian distribution on the natural log scale. The remaining line of equation (6.4) is a prior for σ^2, which is specified as a non-informative conjugate inverse-gamma distribution. This model estimates $\mathbf{x}$ by its marginal posterior distribution

$$f(\mathbf{x}|W(\mathcal{S}),\mathbf{y}) = \int_{\beta,\sigma^2_\epsilon} f(\mathbf{x},\beta,\sigma^2_\epsilon|W(\mathcal{S}),\mathbf{y})\mathrm{d}\beta\mathrm{d}\sigma^2_\epsilon, \tag{6.5}$$

which incorporates the possibility of measurement error in the ambient monitoring data. Estimation of $(\beta,\mathbf{x},\sigma^2_\epsilon)$ can be carried out using MCMC simulation, details of which are given below.

6.3.1 MCMC estimation algorithm

The joint posterior distribution of $(\beta,\ln(\mathbf{x}),\sigma^2_\epsilon)$ is given by

$$\begin{aligned} f(\beta,\ln(\mathbf{x}),\sigma^2_\epsilon|W(\mathbf{s}),\mathbf{y}) \;\propto\; & \prod_{i=1}^{n}\mathrm{Poisson}(y_i|\beta,x_i)\prod_{i=1}^{n}\prod_{l=1}^{k}\mathrm{N}(\ln(w_i(\mathbf{s}_l))|\ln(x_i),\sigma^2_\epsilon) \\ & \times \prod_{i=1}^{n}\mathrm{N}(\ln(x_i)|\mu_x,\sigma^2_x)\mathrm{N}(\beta|\mu_\beta,\Sigma_\beta)\text{Inverse-Gamma}(\sigma^2_\epsilon|e,f), \end{aligned}$$

and the parameters are updated using a block Metropolis-Hastings algorithm, in which starting values $(\beta^{(0)},\ln(\mathbf{x})^{(0)},\sigma^{2(0)}_\epsilon)$ are generated from overdispersed versions of the priors (for example t-distributions replace Gaussian distributions). The algorithm alternately samples from the full conditional distributions of the following blocks.

(a) **Regression parameters $\beta = (\gamma,\alpha)$**

The full conditional of β is identical to the full conditional from a Poisson generalised linear model, and can be sampled from using a Metropolis-Hastings algorithm with a random walk proposal. This algorithm is described in chapter 2.3, and has also been used in chapters four and five.

(b) **Average pollution levels (logged) $\ln(\mathbf{x}) = (\ln(x_1)\ldots,\ln(x_n))$**

The full conditional of $\ln(\mathbf{x})$ is not a standard distribution, but can also be sampled from using a Metropolis-Hastings algorithm. The vector can be updated in blocks, $\ln(\mathbf{x})_{r,s} = (\ln(x_r)\ldots,\ln(x_s))$, whose full conditional is given by

$$f(\ln(\mathbf{x})_{r,s}|\beta,\sigma_\epsilon^2,\mathbf{y},W(\mathcal{S}))\propto\prod_{i=r}^{s}\mathrm{N}(\ln(x_i)|m_i,\tau^2)\prod_{i=r}^{s}\mathrm{Poisson}(y_i|x_i,\beta),$$

where

$$\begin{aligned} m_i &= \left(\frac{1}{\sigma_x^2}+\frac{k}{\sigma_\epsilon^2}\right)^{-1}\left(\frac{\mu_x}{\sigma_x^2}+\frac{1}{\sigma_\epsilon^2}\sum_{l=1}^{k}\ln(w_i(\mathbf{s}_l))\right), \\ \tau^2 &= \left(\frac{1}{\sigma_x^2}+\frac{k}{\sigma_\epsilon^2}\right)^{-1}. \end{aligned}$$

Each block is updated using a random walk proposal, because the acceptance probability simplifies to the ratio of full conditionals and the proposals are cheap to generate.

(c) Measurement error variance σ_ϵ^2

The full conditional of σ_ϵ^2 is an inverse-gamma(e^*,f^*) distribution, where

$$\begin{aligned} e^* &= e+\frac{nk}{2}, \\ f^* &= f+\frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{k}(\ln(\mathbf{w}_i(\mathbf{s}_l))-\ln(x_i))^2, \end{aligned}$$

meaning that it can be updated using a Gibbs sampling step.

6.4 A spatio-temporal model

This model extends that of the previous section by explicitly modelling the spatial variation in the ambient pollution data. As a result, let $X(\mathcal{S})=(\mathbf{x}_1(\mathcal{S})^{\mathrm{T}},\ldots,\mathbf{x}_n(\mathcal{S})^{\mathrm{T}})_{n\times k}$ denote the true (unobserved) pollution levels at sites in $\mathcal{S}$ for all n days. The joint likelihood of these data is given by

$$f(\mathbf{y},\mathbf{x},W(\mathcal{S}),X(\mathcal{S})|\Omega)=f(\mathbf{y}|\mathbf{x},\omega_1)f(\mathbf{x}|W(\mathcal{S}),X(\mathcal{S}),\omega_2)f(W(\mathcal{S}),X(\mathcal{S})|\omega_3), \qquad (6.6)$$

where $\Omega=(\omega_1,\omega_2,\omega_3)$ denotes the vector of parameters. The Bayesian spatio-temporal model combines this likelihood with a prior $f(\Omega)$, and is split into three distinct parts.

(a) Pollution model - The third term in equation (6.6) is the joint distribution of the measured and true pollution levels at the k monitor locations, which is represented by a spatio-temporal pollution model. Such models are the focus of much

current research (for a review see Sahu and Mardia (2005b)), and a general form is given by equation (2.16) with the addition of priors for a Bayesian analysis. The transformation $g()$ is applied to aid normality and stabilise the variance, and common examples include natural log (Shaddick and Wakefield (2002)) or square root (Sahu et al. (2006)). The first line of equation (2.16) is a classical measurement error model, which relates the noisy pollution data to the underlying 'true' unobserved levels. The second line is a spatio-temporal model for these true pollution levels, which comprises a mean function $\mu_i(\mathbf{s}_l|\boldsymbol{\delta})$ and a zero-mean spatio-temporal process $V_i(\mathbf{s}_l|\boldsymbol{\phi})$. Previous studies have modelled the mean function with a trend surface model (Zhu et al. (2003)), cyclical variation (Tonellato (2001)), a temporal only trend (Zidek et al. (2002)) and the Kriged-kalman model (Sahu and Mardia (2005a)), while $V_i(\mathbf{s}_l|\boldsymbol{\phi})$ has taken the form of a multivariate autoregressive process (see Wikle et al. (1998) and Zidek et al. (2002)) and the sum or product of spatial and temporal components (see Shaddick and Wakefield (2002), Sahu and Mardia (2005a) and Zhu et al. (2003)). The covariance function of $V_i(\mathbf{s}_l|\boldsymbol{\phi})$ is typically assumed to be stationary, isotropic and separable (multiplicative or additive), because the number of monitor locations (k) is too small to permit accurate estimates with more complex models. In this chapter I use the model of Shaddick and Wakefield (2002), which is given by

$$
\begin{aligned}
\ln(\mathbf{w}_i(\mathcal{S})) &\sim \mathrm{N}(\ln(\mathbf{x}_i(\mathcal{S})), \sigma_\epsilon^2 \mathrm{I}) \quad \text{for } i = 1, \ldots, n, \\
\ln(\mathbf{x}_i(\mathcal{S})) &= B_i\boldsymbol{\delta} + \mathbf{1}\theta_i + \mathbf{m}(\mathcal{S}), \\
\theta_i &\sim \mathrm{N}(\phi_1\theta_{i-1}, \sigma_\theta^2) \quad \text{for } i = 2, \ldots, n, \\
\mathbf{m}(\mathcal{S}) &\sim \mathrm{N}(\mathbf{0}, \sigma_m^2\Sigma(\phi_2)), \\
\boldsymbol{\delta} &\sim \mathrm{N}(\mu_\delta, \Sigma_\delta), \\
\sigma_\epsilon^2 &\sim \text{Inverse Gamma}(e_1, f_1), \\
\sigma_\theta^2 &\sim \text{Inverse Gamma}(e_2, f_2), \\
\sigma_m^2 &\sim \text{Inverse Gamma}(e_3, f_3), \\
f(\phi_1) &\propto 1, \\
\phi_2 &\sim \text{Uniform}(p_1, p_2),
\end{aligned}
\tag{6.7}
$$

where $\omega_3 = (\boldsymbol{\delta}, \mathbf{m}(\mathcal{S}), \boldsymbol{\theta}, \phi_1, \phi_2, \sigma_m^2, \sigma_\theta^2, \sigma_\epsilon^2)$. The mean function is represented by a trend surface model $B_i\boldsymbol{\delta}$, which has a Gaussian prior with known mean and variance. The spatio-temporal process is the sum of spatial ($\mathbf{m}(\mathcal{S}) = (m(\mathbf{s}_1), \ldots, m(\mathbf{s}_k))$) and temporal ($\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$) components, which forces the covariance function to be additively separable. Temporal correlation is modelled by a first order autoregressive process, where θ_1 is assigned a flat prior and $\mathbf{1}$ denotes a $k \times 1$ vector of ones that

force the process to be identical at each monitor location. The spatial correlation is modelled by a set of zero-mean Gaussian random effects, which are time invariant and have correlation matrix $\Sigma(\phi_2)$. To ensure the model is identifiable, the random effects are constrained to sum to zero as suggested by Shaddick and Wakefield (2002). The spatial correlation matrix is given an exponential structure, with ijth element

$$\Sigma_{ij} = \exp(-\phi_2 ||\mathbf{s}_i - \mathbf{s}_j||), \quad . \tag{6.8}$$

where $||.||$ denotes the Euclidean norm. The strength of the correlation is controlled by ϕ_2, which is assigned a uniform prior that forces the correlation at the maximum distance in $\mathcal{R}$ to be between 0.05 and 0.95. In common with the measurement error model the variance parameters are assigned conjugate inverse-gamma priors, whose hyperparameters are chosen so that the distribution is non-informative.

(b) Spatial aggregation model - The second term in equation (6.6) simplifies to $f(\mathbf{x}|W(\mathcal{S}), \omega_3)$, because $X(\mathcal{S})$ is completely specified by ω_3. This distribution has no closed form expression, meaning that it does not have any additional parameters (ω_2 does not exist) and only depends on the spatio-temporal pollution model through ω_3. It can be estimated by simulation using a variety of techniques, including spatial imputation (Duddek et al. (1995)), ordinary and universal kriging (Carlin et al. (1999)) and the point-block realignment approach of Gelfand et al. (2001) (Zhu et al. (2003)). The last of these derives a closed form expression for this predictive distribution, under the simplifying assumptions that there is no measurement error and both $\mathbf{x}$ and $W(\mathcal{S})$ are multivariate Gaussian. The work of Duddek et al. (1995) and Carlin et al. (1999) are based on a two stage strategy, which estimates the average pollution levels using a spatio-temporal model and plugs these values into a mortality model. In contrast Zhu et al. (2003) propose a combined approach, that simultaneously estimates the average pollution level and its corresponding relationship with mortality. However they report that this combined model is computationally impractical to implement using MCMC methods, and propose a simplification that separates the simulation of the health and pollution models. This technique is known as cutting 'feedback', and simplifies the full conditional distribution of $\mathbf{x}$ by removing its dependence on the mortality data. In this chapter $\mathbf{x}$ is estimated using an approach similar to that adopted by Zhu et al. (2003), which is based on Bayesian spatial prediction. Estimation is implemented using a single MCMC algorithm, but the feedback between the pollution and mortality models is cut for computational efficiency. As a result the simulation of $\mathbf{x}$ does not depend on the mortality data, meaning that it is estimated by the predictive distribution

$$f(\mathbf{x}|W(\mathcal{S})) = \int_{\omega_3} f(\mathbf{x}|\omega_3) f(\omega_3|W(\mathcal{S})) \mathrm{d}\omega_3. \tag{6.9}$$

Full details of the MCMC simulation algorithm are given below.

(c) **Mortality model** - The mortality model is identical to equation (3.2), except that the surrogate exposure $\mathbf{w}$ is replaced by the true exposure $\mathbf{x}$.

6.4.1 MCMC estimation algorithm

The joint posterior distribution is estimated using a block Metropolis-Hastings algorithm, in which starting values $(\boldsymbol{\beta}^{(0)}, \mathbf{x}^{(0)}, \sigma_\epsilon^{2(0)}, \boldsymbol{\delta}^{(0)}, \boldsymbol{\theta}^{(0)}, \mathbf{m}(\mathcal{S})^{(0)}, \phi_1^{(0)}, \phi_2^{(0)}, \sigma_\theta^{2(0)}, \sigma_m^{2(0)})$ are generated from overdispersed versions of the priors (for example t-distributions replace Gaussian distributions). A full MCMC algorithm is computationally prohibitive, so the short cut used by Zhu et al. (2003) is adopted here. This short cut separates the simulation of $\boldsymbol{\beta}$ from that of $(\mathbf{x}, \sigma_\epsilon^2, \boldsymbol{\delta}, \boldsymbol{\theta}, \mathbf{m}(\mathcal{S}), \phi_1, \phi_2, \sigma_\theta^2, \sigma_m^2)$, which cuts the feedback between the mortality and pollution models. Although this is not strictly Bayesian, the average pollution level is unlikely to depend on the counts of mortality, meaning that cutting feedback should not be overly restrictive. The algorithm is implemented by updating the parameters in three batches.

(a) **Spatio-temporal pollution model** $(\sigma_\epsilon^2, \boldsymbol{\delta}, \boldsymbol{\theta}, \mathbf{m}(\mathcal{S}), \phi_1, \phi_2, \sigma_\theta^2, \sigma_m^2)$

The spatial parameters are updated in blocks using a Gibbs sampling algorithm, further details of which are given below.

(b) **Spatial aggregation model** $f(\mathbf{x}|W(\mathcal{S}), \omega_3)$

The conditional distribution of $\mathbf{x}$ is non-standard, and its simulation is discussed below.

(c) **Mortality model** $\boldsymbol{\beta} = (\boldsymbol{\gamma}, \boldsymbol{\alpha})$

The full conditional of $\boldsymbol{\beta}$ is identical to the full conditional from the standard Poisson model (3.2), and is updated using a block Metropolis-Hastings algorithm with a random walk proposal.

Details of steps (a) and (b) are given below.

(a) Spatio-temporal pollution model

The parameters are Gibbs sampled in the following blocks

1. $f(\boldsymbol{\delta}|W(\mathcal{S}), \mathbf{m}(\mathcal{S}), \boldsymbol{\theta}, \sigma_\epsilon^2) \sim \mathrm{N}(\boldsymbol{\mu}_*, \Sigma_*)$ where

$$\begin{aligned}
\mu_* &= \left(\Sigma_\delta^{-1} + \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n} B_i^{\mathrm{T}} B_i\right)^{-1} \left(\Sigma_\delta^{-1}\mu_\delta + \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n} B_i^{\mathrm{T}}(\ln(\mathbf{w}_i(\mathcal{S})) - \mathbf{1}\theta_i - \mathbf{m}(\mathcal{S}))\right) \\
\Sigma_* &= \left(\Sigma_\delta^{-1} + \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n} B_i^{\mathrm{T}} B_i\right)^{-1}
\end{aligned}$$

2. $f(\sigma_\epsilon^2 | W(\mathcal{S}), \delta, \mathbf{m}(\mathcal{S}), \theta) \sim$ Inverse Gamma(e_*, f_*) where

$$\begin{aligned}
e_* &= e_1 + \frac{nk}{2} \\
f_* &= f_1 + \frac{1}{2}\sum_{i=1}^{n} (\ln(\mathbf{w}_i(\mathcal{S})) - B_i\delta - \mathbf{1}\theta_i - \mathbf{m}(\mathcal{S}))^{\mathrm{T}} (\ln(\mathbf{w}_i(\mathcal{S})) - B_i\delta - \mathbf{1}\theta_i - \mathbf{m}(\mathcal{S}))
\end{aligned}$$

3. $f(\mathbf{m}(\mathcal{S}) | W(\mathcal{S}), \delta, \theta, \sigma_\epsilon^2, \sigma_m^2, \phi_2) \sim \mathrm{N}(\mu_*, \Sigma_*)$ where

$$\begin{aligned}
\mu_* &= \left(\frac{1}{\sigma_m^2}\Sigma(\phi_2)^{-1} + \frac{n}{\sigma_\epsilon^2}\mathrm{I}\right)^{-1} \left(\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}(\ln(\mathbf{w}_i(\mathcal{S})) - B_i\delta - \mathbf{1}\theta_i)\right) \\
\Sigma_* &= \left(\frac{1}{\sigma_m^2}\Sigma(\phi_2)^{-1} + \frac{n}{\sigma_\epsilon^2}\mathrm{I}\right)^{-1}
\end{aligned}$$

4. $f(\sigma_m^2 | \mathbf{m}(\mathcal{S}), \phi_2) \sim$ Inverse Gamma(e_*, f_*) where

$$\begin{aligned}
e_* &= e_3 + \frac{k}{2} \\
f_* &= f_3 + \frac{1}{2}\mathbf{m}(\mathcal{S})^{\mathrm{T}}\Sigma(\phi_2)^{-1}\mathbf{m}(\mathcal{S})
\end{aligned}$$

5. The full conditional distribution of ϕ_2 is non-standard, and is sampled from using a Metropolis-Hastings algorithm with a random walk proposal.

6. $f(\sigma_\theta^2 | \theta, \phi_1) \sim$ Inverse Gamma(e_*, f_*) where

$$\begin{aligned}
e_* &= e_2 + \frac{n-1}{2} \\
f_* &= f_2 + \frac{1}{2}\sum_{i=2}^{n}(\theta_i - \phi_1\theta_{i-1})^2
\end{aligned}$$

7. $f(\phi_1 | \sigma_\theta^2, \theta) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$$\begin{aligned}\mu_* &= \frac{\sum_{i=2}^n \theta_i\theta_{i-1}}{\sum_{i=2}^n \theta_{i-1}^2} \\ \sigma_*^2 &= \frac{\sigma_\theta^2}{\sum_{i=2}^n \theta_{i-1}^2}\end{aligned}$$

8. The full conditional of $\boldsymbol{\theta}$ cannot be Gibbs sampled in a single block, because its prior distribution has a singular variance matrix. This distribution is given by $f(\boldsymbol{\theta}|\sigma_\theta^2, \phi) = \prod_{i=2}^n \mathrm{N}(\theta_i|\phi_1\theta_{i-1}, \sigma^2) \propto \exp\left(-\frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}} K \boldsymbol{\theta}\right)$, where K has a bandwidth of one and has non-zero elements

$$\begin{aligned}K_{ii} &= \begin{cases} \frac{\phi_1^2}{\sigma_\theta^2} & i = 1 \\ \frac{\phi_1^2+1}{\sigma_\theta^2} & i = 1, \ldots, n-1 \\ \frac{1}{\sigma_\theta^2} & i = n \end{cases}, \\ K_{i,i+1} &= -\frac{\phi_1}{\sigma_\theta^2} \qquad \forall\, i, \\ K_{i,i-1} &= -\frac{\phi_1}{\sigma_\theta^2} \qquad \forall\, i.\end{aligned}$$

However, partitioning $\boldsymbol{\theta}$ into blocks simplifies the prior into a set of standard Gaussian distributions, allowing Gibbs sampling steps to be used. For a block $\boldsymbol{\theta}_{r,s} = (\theta_r, \ldots, \theta_s)$ the prior is given by $\boldsymbol{\theta}_{rs}|\boldsymbol{\theta}_{-rs} \sim \mathrm{N}(\boldsymbol{\mu}_{rs}, \Sigma_{rs})$, which has mean and variance given by

$$\Sigma_{rs} = \begin{pmatrix} \frac{\phi_1^2+1}{\sigma_\theta^2} & -\frac{\phi_1}{\sigma_\theta^2} & & & \\ -\frac{\phi_1}{\sigma_\theta^2} & \frac{\phi^2+1}{\sigma_\theta^2} & & & \\ & \ddots & \ddots & \ddots & \\ & & & \frac{\phi_1^2+1}{\sigma_\theta^2} & -\frac{\phi_1}{\sigma_\theta^2} \\ & & & -\frac{\phi_1}{\sigma_\theta^2} & \frac{\phi_1^2+1}{\sigma_\theta^2} \end{pmatrix}^{-1} \quad \text{and}$$

$$\boldsymbol{\mu}_{rs} = \Sigma_{rs} \begin{bmatrix} \frac{\phi_1\theta_{r-1}}{\sigma_\theta^2} \\ 0 \\ \vdots \\ 0 \\ \frac{\phi_1\theta_{s+1}}{\sigma_\theta^2} \end{bmatrix}$$

respectively. However the mean and variance are slightly different if $r = 1$ or $s = n$. The data likelihood for this block is given by $\boldsymbol{\theta}_{rs}|W(\mathcal{S}) \sim \mathrm{N}(\boldsymbol{\gamma}_{rs}, \Omega_{rs})$ which has

mean and variance given by

$$\begin{aligned}\Omega_{rs} &= \frac{\sigma_\epsilon^2}{k}\mathrm{I} \text{ and} \\ \gamma_{rs} &= \begin{pmatrix} \frac{1}{k}\sum_{l=1}^{k}\ln(w_r(\mathbf{s}_l)) - B_{r,\mathbf{s}_l}\boldsymbol{\delta} - m(\mathbf{s}_l) \\ \vdots \\ \frac{1}{k}\sum_{l=1}^{k}\ln(w_s(\mathbf{s}_l)) - B_{s,\mathbf{s}_l}\boldsymbol{\alpha} - m(\mathbf{s}_l) \end{pmatrix}\end{aligned}$$

respectively. Therefore the full conditional of $\boldsymbol{\theta}_{rs}$ is given by

$$\boldsymbol{\theta}_{rs} \sim \mathrm{N}\left([\Omega_{rs}^{-1} + \Sigma_{rs}^{-1}]^{-1}[\Omega_{rs}^{-1}\gamma_{rs} + \Sigma_{rs}^{-1}\boldsymbol{\mu}_{rs}]\ ,\ [\Omega_{rs}^{-1} + \Sigma_{rs}^{-1}]^{-1}\right),$$

and the entire vector can be updated in a small number of blocks.

(b) Spatial aggregation model

The spatially averaged pollution levels are sampled using a three stage process, similar to that adopted by Zhu et al. (2003). At iteration j, $\mathbf{x}^{(j)}$ is generated as follows.

1. Generate c random locations $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_c\}$ across $\mathcal{R}$.

2. Predict the true pollution level at each of these c locations by sampling from the predictive distribution $f(X(\mathcal{V})|\boldsymbol{\omega}_3^{(j)})$. For day i a sample is generated by calculating

$$\mathbf{x}_i(\mathcal{V})^{(j)} = \exp(B_{i,\mathcal{V}}\boldsymbol{\delta}^{(j)} + \mathbf{1}\theta_i^{(j)} + \mathbf{m}(\mathcal{V})^{(j)}) \quad \text{for } i = 1, \ldots, n,$$

   where $B_{i,\mathcal{V}}$ is the matrix of covariates for the trend model, while $\mathbf{m}(\mathcal{V})^{(j)}$ is the spatial random effects at the c prediction locations. The latter are random samples from the conditional distribution $f(\mathbf{m}(\mathcal{V})|\mathbf{m}(\mathcal{S})^{(j)})$, which is multivariate Gaussian with

$$\begin{aligned}\mathbb{E}[\mathbf{m}(\mathcal{V})|\mathbf{m}(\mathcal{S})^{(j)}] &= \Sigma(\phi_2^{(j)})_{\mathcal{V}\mathcal{S}}\Sigma(\phi_2^{(j)})_{\mathcal{S}}^{-1}\mathbf{m}(\mathcal{S})^{(j)} \text{ and} \\ \mathrm{Var}[\mathbf{m}(\mathcal{V})|\mathbf{m}(\mathcal{S})^{(j)}] &= \sigma_m^{2(j)}[\Sigma(\phi_2^{(j)})_{\mathcal{V}} - \Sigma(\phi_2^{(j)})_{\mathcal{V}\mathcal{S}}\Sigma(\phi_2^{(j)})_{\mathcal{S}}^{-1}\Sigma(\phi_2^{(j)})_{\mathcal{V}\mathcal{S}}^{\mathrm{T}}].\end{aligned}$$

   Here $(\Sigma(\phi_2^{(j)})_{\mathcal{S}}, \Sigma(\phi_2^{(j)})_{\mathcal{V}}, \Sigma(\phi_2^{(j)})_{\mathcal{V}\mathcal{S}})$ denote the variance and covariance matrices for the spatial sites $(\mathcal{S}, \mathcal{V})$, and are calculated using the exponential correlation model given by equation (6.8).

3. Set each $x_i^{(j)}$ as the average of the predictions:

Table 6.1: Summary of the simulated pollution data.

Scenario	Measurement error	Spatial variation		Number of Monitors
		variance	Correlation	
1	low $\sigma_\epsilon^2 = 0.00007$	low $\sigma_m^2 = 0.00007$	low $\phi_2 = 0.0857$	low 5
2	low $\sigma_\epsilon^2 = 0.00007$	low $\sigma_m^2 = 0.00007$	low $\phi_2 = 0.0857$	high 20
3	low $\sigma_\epsilon^2 = 0.00007$	low $\sigma_m^2 = 0.00007$	high $\phi_2 = 0.0133$	low 5
4	low $\sigma_\epsilon^2 = 0.00007$	low $\sigma_m^2 = 0.00007$	high $\phi_2 = 0.0133$	high 20
5	low $\sigma_\epsilon^2 = 0.00007$	high $\sigma_m^2 = 0.0289$	low $\phi_2 = 0.0857$	low 5
6	low $\sigma_\epsilon^2 = 0.00007$	high $\sigma_m^2 = 0.0289$	low $\phi_2 = 0.0857$	high 20
7	low $\sigma_\epsilon^2 = 0.00007$	high $\sigma_m^2 = 0.0289$	high $\phi_2 = 0.0133$	low 5
8	low $\sigma_\epsilon^2 = 0.00007$	high $\sigma_m^2 = 0.0289$	high $\phi_2 = 0.0133$	high 20
9	high $\sigma_\epsilon^2 = 0.0289$	low $\sigma_m^2 = 0.00007$	low $\phi_2 = 0.0857$	low 5
10	high $\sigma_\epsilon^2 = 0.0289$	low $\sigma_m^2 = 0.00007$	low $\phi_2 = 0.0857$	high 20
11	high $\sigma_\epsilon^2 = 0.0289$	low $\sigma_m^2 = 0.00007$	high $\phi_2 = 0.0133$	low 5
12	high $\sigma_\epsilon^2 = 0.0289$	low $\sigma_m^2 = 0.00007$	high $\phi_2 = 0.0133$	high 20
13	high $\sigma_\epsilon^2 = 0.0289$	high $\sigma_m^2 = 0.0289$	low $\phi_2 = 0.0857$	low 5
14	high $\sigma_\epsilon^2 = 0.0289$	high $\sigma_m^2 = 0.0289$	low $\phi_2 = 0.0857$	high 20
15	high $\sigma_\epsilon^2 = 0.0289$	high $\sigma_m^2 = 0.0289$	high $\phi_2 = 0.0133$	low 5
16	high $\sigma_\epsilon^2 = 0.0289$	high $\sigma_m^2 = 0.0289$	high $\phi_2 = 0.0133$	high 20

$$x_i^{(j)} = \frac{1}{c}\sum_{l=1}^{c} x_i(\mathbf{v}_l)^{(j)}.$$

For computational efficiency the same set of prediction locations will be used at each iteration, while the choice of c requires a tradeoff to be made. As c increases estimates of $\mathbf{x}$ will become more accurate, but at the expense of a larger computational burden.

6.5 Simulation study

This section describes a simulation study which compares the performance of the standard (equation (3.2)), measurement error (equation (6.4)) and spatio-temporal (equation (6.7)) models. These models are applied to numerous sets of pollution data that have three defining characteristics: (i) the amount of spatial variation in the underlying pollution surface; (ii) the proportion of measurement error in the observed pollution data; (iii) the number of pollution monitors that are available. The study is motivated by the three questions posed in section 6.2, and is split into three subsections. The first describes how the pollution and mortality data are generated, the second investigates questions (i) and (ii), while the third focuses on question (iii).

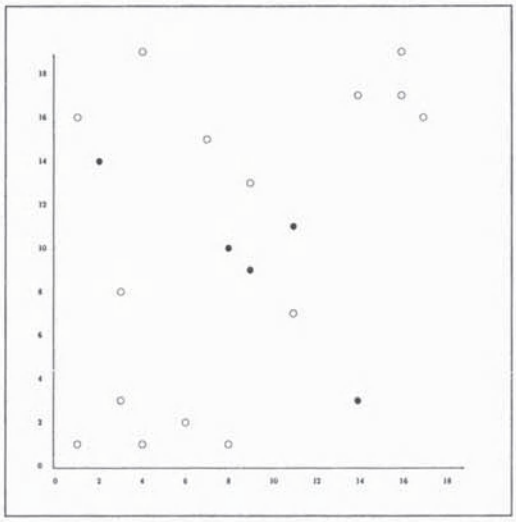
6.5.1 Description of the data

All three models are applied to sixteen sets of pollution and mortality data, which are summarised in Table 6.1. Each pollution surface has four defining parameters: spatial variation (σ_m^2), spatial correlation (ϕ_2), measurement error (σ_ϵ^2), and the number of monitors (k), each of which can be either high or low. The pollution surfaces are simulated at 400 locations on a 20km $\times$ 20km grid of points ($(0,0)$ up to $(19,19)$ at intervals of one km), for 365 consecutive days. The surfaces are generated using model (6.7), and have the same trend surface model ($B_i\delta$) and temporal correlation (θ) in all cases. This ensures that the only differences between each scenario are those described in Table 6.1, which prevents the comparisons being contaminated by other factors. The trend surface model comprises an intercept term, quadratic trends in longitude and latitude, and three pairs of interactions between longitude, latitude and time, which are used so that the pollution surface is not completely flat. However this trend is chosen to be relatively small, so that the differences in (σ_m^2, ϕ_2) between each scenario are not hidden. The intercept term is fixed at 3.4 (approximately 30 on the original scale) so that the mean pollution level is similar to the NO_2 and O_3 data analysed in section 6.6. Temporal variation and correlation are induced into the pollution data by a first order autoregressive process, which is initialised by setting $\theta_0 = 0$. The variance and correlation are fixed at $\phi_1 = 0.7$ and $\sigma_\theta^2 = 0.0625$ respectively, which ensures the temporal characteristics of the simulated data are similar to those analysed in section 6.6.

The spatial characteristics of the pollution data are controlled by σ_m^2 and ϕ_2, which determine their variance and correlation respectively. The variance is chosen so that the likely range of spatial variation is 1% (low, $\sigma_m^2 = 0.00007$) or 20% (high, $\sigma_m^2 = 0.0289$) of the mean pollution level (3.4 on the logged scale), which is achieved by equating $\pm 2\sigma_m$ (the likely range of values) to 1% or 20% of 3.4. The correlation is chosen so that it decays to 0.1 (low, $\phi_2 = 0.0857$) or 0.7 (high, $\phi_2 = 0.0133$) at the maximum distance in the grid (26.87), which represents the range of values found in real pollution data. The amount of measurement error (controlled by σ_ϵ^2) is also fixed at either 1% or 20% of the mean pollution level, and the equality with σ_m^2 allows their relative effects to be compared fairly. The possible combinations of these dichotomies result in eight unique pollution fields, each of which is analysed assuming a small (5) or large (20) number of spatial observations are available. This corresponds to the number of pollution monitors typically available in a single city, and their locations are shown in Figure 6-1. The first five are fixed to be similar to the grouping found in the Greater London data analysed in section 6.6, while the remaining fifteen are selected at random, a point which is taken up chapter seven.

The simulated mortality data are generated using a Poisson regression model, similar to that given in equation (3.2). The spatially averaged pollution levels ($\mathbf{x}$) are calculated

Figure 6-1: The locations of the ambient monitors in the 20km × 20km grid of simulated values. The filled circles represent the monitors that are used in the odd numbered scenarios, while all locations are used in the remaining scenarios.



by averaging the values at all 400 locations on the grid, while the covariates ($\mathbf{z}_i$) include an intercept term, a natural cubic spline of temperature with three degrees of freedom and sinusoidal components with periods of a whole, half and a quarter of a year. The intercept term is fixed at 3.4 (approximately 30 on the original scale), which ensures the daily mortality counts are similar to those observed in Greater London. The association between air pollution and mortality is fixed at an increase in risk of 2% for an increase in ten units of pollution, which is of a similar magnitude to that found in previous studies (see Dockery and Pope (1994) and Samet et al. (2000)). To ensure that only the levels of spatial variation and measurement error affect the results, the same set of covariates ($\mathbf{z}_i, B_i$) that generated the pollution and mortality data are used in the final models.

6.5.2 Assessment of the standard model

The standard model (given by equation (3.2)) was initially applied to 200 simulated data sets from each scenario, which ensures the results are not affected by a single realisation of the pollution surface. However this results in 3,200 separate analyses, so for computational ease a likelihood approach to estimation was adopted. The estimated relative risks for all 3,200 data sets are shown in Figure 6-2, and are summarised for each scenario in Table 6.2. These summaries include the mean and standard deviation of the absolute bias for each scenario, which are calculated as

$$\mu_{\text{bias}} = \frac{1}{200}\sum_{i=1}^{200} |\hat{RR}(\gamma^{(i)}) - 1.02| \text{ and } \text{sd}_{\text{bias}} = \left[\frac{1}{199}\sum_{i=1}^{200}(|\hat{RR}(\gamma^{(i)}) - 1.02| - \mu_{\text{bias}})^2\right]^{1/2},$$

where $\hat{RR}(\gamma^{(i)})$ denotes the estimated relative risk for the ith data set. All estimated relative risks are less than 1% away from their true value (between (1.01, 1.03)), suggesting that the standard model produces results of the correct overall magnitude regardless of the underlying pollution surface. The standard model performs best if the pollution surface is spatially flat and exhibits minimal measurement error (scenarios one to four), which can be seen by examining μ_{bias} for each scenario. In this situation the mean absolute bias ranges from 0.008% to 0.01% on the relative risk scale, suggesting that the standard model is very accurate if neither spatial variation nor measurement error are present. The results from the remaining scenarios are much more varied, and the bias in relative risk for a single data set is unpredictable. This unpredictability is probably caused by the locations of the monitoring sites in relation to the peaks and troughs in the observed pollution surface, meaning that only the average impact of spatial variation, measurement error and the number of monitors can be assessed.

(a) - Spatial variation only

If the pollution data exhibit substantial spatial variation (scenarios five and six which have high variation and low correlation) the mean absolute bias increases by between five and nine times. For example, in scenario five the mean absolute bias is approximately 0.09% which compares with 0.01% for scenario one. This bias reduces if the correlation is high (scenarios seven and eight), being between two and five times larger than when the variation is low. This indicates that changes in both spatial variance and correlation affect the standard model, although a comparison of scenarios two (low - σ_m^2, low - ϕ_2) and eight (high - σ_m^2, high - ϕ_2) show that high spatial variance has twice the impact of low correlation

(b) - Measurement error only

If the pollution data exhibit substantial measurement error (scenarios nine to twelve) the bias rises by between eight and nineteen times. For example, in scenario nine the mean absolute bias is approximately 0.19% compared to 0.01% for scenario one. This increase in bias is approximately twice as large as that observed for the increase in spatial variation, suggesting that measurement error has a larger detrimental impact on the standard model. In addition, Figure 6-2 shows that the presence of measurement error biases the estimated relative risks towards one (no effect) on average, where as spatial variation has no such effect.

Table 6.2: Summary of the absolute bias in relative risks from using the standard model.

Scenario	Mean (μ_{bias})	Standard deviation (sd_{bias})
1	0.00010	0.00008
2	0.00008	0.00005
3	0.00010	0.00008
4	0.00008	0.00005
5	0.00088	0.00069
6	0.00047	0.00036
7	0.00047	0.00032
8	0.00021	0.00016
9	0.00195	0.00149
10	0.00087	0.00063
11	0.00203	0.00150
12	0.00096	0.00074
13	0.00193	0.00141
14	0.00094	0.00074
15	0.00206	0.00152
16	0.00097	0.00072

(c) - Spatial variation and measurement error

If the pollution data are contaminated with measurement error the addition of spatial variation has little impact on the performance of the standard model. For example the mean absolute bias is 0.203% (scenario eleven) if only measurement error is present, which compares with 0.0206% (scenario fifteen) if spatial variation is added. In contrast, a comparison of scenarios six and fourteen show that adding measurement error to spatially variable pollution data doubles the mean absolute bias.

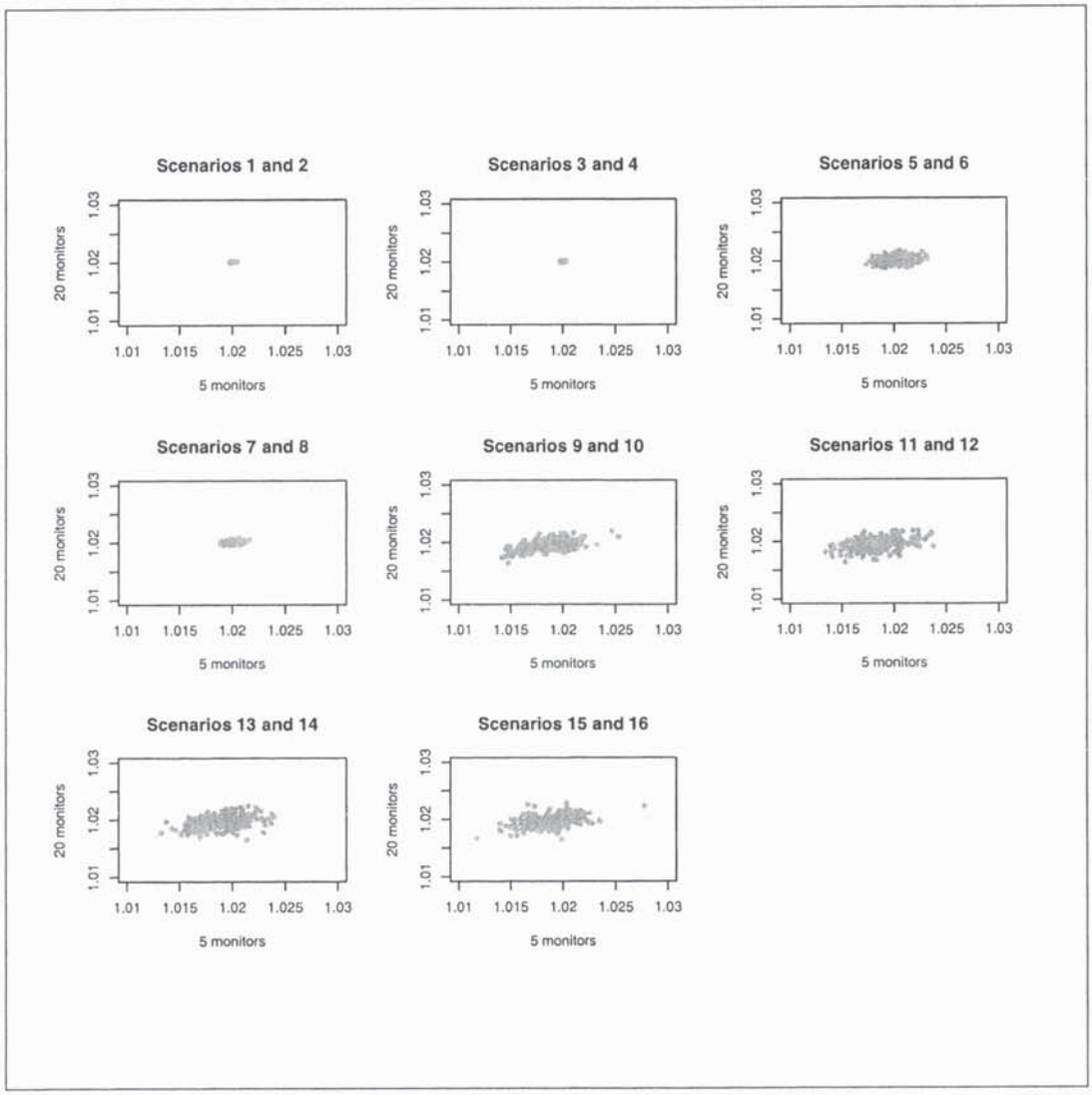
(d) - Number of monitors

If the pollution surface is spatially flat with no measurement error (scenarios one to four) the number of monitors does not affect the accuracy of the standard model (for example average biases of 0.001% and 0.0008% from scenarios one and two). However if either of these factors are present, the mean and standard deviation of the absolute bias doubles if only five monitors are present compared with the full set of twenty.

6.5.3 Comparison of the standard, measurement error and spatio-temporal models

To compare the performance of the standard, measurement error and spatio-temporal models, they are applied to a single data set from each of the sixteen scenarios. In each scenario the data set with the largest absolute bias (as estimated from the standard

Figure 6-2: Estimated relative risks from the standard model (equation (see 3.2)) for each of the sixteen scenarios.



model) is chosen, because the aim is to determine whether the more complex models produce better results. The models are implemented using the MCMC simulation algorithms previously described, and posterior inference is based on 16,000 samples. In each case a Markov chain is burnt in for 40,000 iterations (by which point convergence had been reached), and run for an additional 80,000 iterations which are thinned by five to reduce the correlation between samples. For the measurement error and spatio-temporal models, sensitivity analyses were conducted to determine whether the results are dependent on the prior distributions. These analyses focused on the variance and correlation parameters, because the other parameters are given standard priors that are known to work well in a variety of situations. A series of 'non-informative' inverse-gamma(ϵ, ϵ) priors with $\epsilon = 1,\ 0.1,\ 0.01,\ 0.001$ were applied to each of the variance parameters, as well as a flat prior on the standard deviation scale. However the estimated variance parameters in this chapter are larger than those in chapter five, and their values showed no dependence on the choice of inverse-gamma prior. As a result inverse-gamma$(0.01, 0.01)$ priors are used in all cases. For the spatio-temporal model a number of uniform priors for ϕ_2 were investigated, and the choice of (p_1, p_2) had no impact on the estimated relative risks. However the estimates of ϕ_2 did not converge for any of the sixteen scenarios, and 95% credible intervals were close to the prior limits. This phenomenon was observed even when a larger number of spatial data points were used (up to 100), and the reasons for this are discussed further in chapter seven. Two possible solutions are to use an informative prior distribution or estimate ϕ_2 prior to the MCMC algorithm, neither of which are ideal. As a result I adopt prior limits that allow the correlation to take on a wide range of reasonable values, which I have found has little or no impact on the corresponding relative risks.

The relative performance of the three models are summarised by Table 6.3, which presents the estimated relative risks for all sixteen scenarios. In addition Table 6.4 summarises how well each model estimates the true exposure $\mathbf{x}$, which is measured as the mean absolute difference between the true and estimated pollution levels (that is $(1/n)\sum_{i=1}^{n}|x_i - \hat{x}_i|$, where $\hat{x}_i$ is the estimated spatially averaged pollution level for day i). In scenarios one to four where spatial variation and measurement error are low, all three models estimate the average pollution level and corresponding relative risk well, with none being preferable in general. In all cases the relative risks are biased by at most 0.04%, while $\mathbf{x}$ is mis-estimated by at most 0.18. The remaining scenarios are discussed below, with particular focus on the impact of spatial variation, measurement error and the number of monitors.

(a) - Spatial variation only

If the pollution data exhibit substantial spatial variation but no measurement error (scenarios five to eight), the spatio-temporal model outperforms both the standard and measurement error approaches. For example, in scenarios five and six the bias in relative risk

is 0.26% (scenario five) and 0.17% (scenario six) for the standard model compared with 0.16% and 0.02% for the spatio-temporal model. The same improvement is observed in the estimated pollution levels, as the mean absolute bias reduces from 2.73 and 1.63 for the standard model to 1.98 and 0.34 for the spatio-temporal model.

(b) - Measurement error only

If the pollution data exhibit substantial measurement error but no spatial variation (scenarios nine to twelve), the spatio-temporal and measurement error models produce similar results and it is difficult to determine which is preferable. Both approaches produce estimates of $\mathbf{x}$ and γ that are slightly less biased than those from the standard model, although the differences are not large. For example, in scenario ten the relative risks are biased by 0.30% (standard), 0.27% (measurement error) and 0.027% (spatio-temporal), while the corresponding mean absolute bias in $\mathbf{x}$ is 1.00, 0.91 and 0.92. The improvement from using the spatio-temporal model is much less pronounced in this situation, suggesting that measurement error is harder to model than spatial variation, a point which is expanded upon in the discussion.

(c) - Spatial variation and measurement error

If spatial variation and measurement error are both present (scenarios thirteen to sixteen), the spatio-temporal model outperforms both alternatives, while the measurement error model outperforms the standard approach. For example, in scenario fourteen the relative risks are biased by 0.26% (standard), 0.24% (measurement error) and 0.17% (spatio-temporal), while the corresponding mean absolute difference in $\mathbf{x}$ is 1.13, 0.98 and 0.88. These differences are larger than those observed for measurement error only but smaller than those seen for spatial variation only, which supports the hypothesis that spatial variation is easier to model.

(d) - Number of monitors

The impact of altering the number of monitors depends on the amount of measurement error and spatial variation present in the pollution data, and has minimal affect if neither factor is present. This can be seen by comparing the results from scenarios three (five monitors) and four (twenty monitors), whose relative risks are biased by 0.04% and 0.02% (standard), 0.01% and 0.02% (measurement error), and 0.03% and 0.01% (spatio-temporal). In contrast, if either factor is present the results improve dramatically as the number of monitors increase. For example, comparing scenarios thirteen and fourteen the bias in relative risk decreases from 0.67% to 0.26% (standard), 0.69% to 0.24% (measurement error) and 0.63% to 0.17% (spatio-temporal), when the number of monitors increases from five to twenty. In addition the mean absolute bias in $\mathbf{x}$ also decreases,

Table 6.3: Estimated relative risks and 95% credible intervals for each data set and model (the true relative risk is 1.02).

Data set	Model		
	Standard	Measurement error	Spatio-temporal
1	1.0203 (1.0038, 1.0370)	1.0201 (1.0039, 1.0370)	1.0202 (1.0037, 1.0365)
2	1.0202 (1.0036, 1.0371)	1.0203 (1.0037, 1.0371)	1.0199 (1.0040, 1.0368)
3	1.0204 (1.0036, 1.0369)	1.0201 (1.0038, 1.0368)	1.0203 (1.0039, 1.0367)
4	1.0202 (1.0039, 1.0370)	1.0202 (1.0038, 1.0367)	1.0199 (1.0036, 1.0364)
5	1.0226 (0.9997, 1.0457)	1.0227 (0.9997, 1.0460)	1.0216 (0.9995, 1.0451)
6	1.0217 (0.9999, 1.0441)	1.0217 (0.9998, 1.0439)	1.0202 (0.9998, 1.0415)
7	1.0214 (1.0061, 1.0373)	1.0215 (1.0061, 1.0372)	1.0208 (1.0057, 1.0366)
8	1.0207 (1.0059, 1.0355)	1.0207 (1.0063, 1.0352)	1.0198 (1.0055, 1.0342)
9	1.0142 (0.9988, 1.0304)	1.0142 (0.9981, 1.0304)	1.0150 (0.9984, 1.0316)
10	1.0170 (1.0010, 1.0336)	1.0173 (1.0007, 1.0337)	1.0173 (1.0007, 1.0344)
11	1.0141 (0.9994, 1.0292)	1.0144 (0.9990, 1.0298)	1.0140 (0.9984, 1.0299)
12	1.0171 (1.0013, 1.0333)	1.0174 (1.0015, 1.0336)	1.0173 (1.0014, 1.0342)
13	1.0133 (0.9943, 1.0325)	1.0131 (0.9942, 1.0326)	1.0137 (0.9947, 1.0337)
14	1.0174 (1.0006, 1.0353	1.0176 (1.0002, 1.0353)	1.0183 (1.0002, 1.0366)
15	1.0116 (0.9945, 1.0290)	1.0127 (0.9949, 1.0301)	1.0138 (0.9959, 1.0321)
16	1.0162 (0.9994, 0.1034)	1.0176 (1.0001, 1.0354)	1.0176 (0.9992, 1.0360)

dropping from 2.44 to 1.13 (standard), 2.67 to 0.98 (measurement error) and 2.30 to 0.88 (spatio-temporal).

6.6 Case study

This section describes a case study which assesses the relative performance of each model by analysing data from Greater London. The first subsection describes the data used in this case study, the second discusses the choice of statistical models, while the third presents the results.

6.6.1 Description of the data

The data used in this case study comprise daily observations from the Greater London area for 1997. The short time frame of one year is adopted for two main reasons, the lack of spatial pollution data for a longer period, and the computational burden of running the spatio-temporal model. The health data comprise daily counts of respiratory mortality within the population living in Greater London, and are part of the data set analysed in chapter four. The pollution data comprise CO, NO_2, O_3 and PM_{10} levels measured at a small number of monitoring sites across Greater London, and their characteristics are summarised in Table 6.5, Figure 6-3 and Figure 6-4. The number of daily pollution

Table 6.4: Mean absolute difference between $\mathbf{x}$ and the estimated values from each model.

Scenario	Standard	Measurement error	Spatio-temporal
1	0.09	0.09	0.18
2	0.05	0.05	0.08
3	0.09	0.09	0.15
4	0.07	0.08	0.05
5	2.73	2.96	1.98
6	1.63	2.00	0.34
7	2.63	2.79	1.69
8	0.82	0.92	0.29
9	1.99	1.95	1.87
10	1.00	0.91	0.92
11	1.95	1.88	1.75
12	0.99	0.93	0.91
13	2.44	2.67	2.30
14	1.13	0.98	0.88
15	1.66	1.67	1.58
16	1.26	0.97	0.85

observations range from 6 to 16, which is similar to the number used in the simulation study. For each pollutant a small fraction (between 3% and 6%) of these observations are missing, and the methods used to overcome this are described in the next subsection. The mean pollution levels for NO_2 (27.9), O_3 (28.1) and PM_{10} (22.8) are very similar, while the underlying level for CO (0.87) is much lower. The amount of spatial and temporal variation in each pollution series is summarised by relative standard deviation (RSD), which is a standardised measure of the coefficient of variation on the percentage scale. The relative standard deviation is calculated as RSD=$|\sigma/\mu| \times 100$, where (μ, σ) denote the sample mean and standard deviation respectively. The spatial relative standard deviation is calculated separately on each day, using the pollution observations at all available monitoring sites for that given day. The corresponding temporal quantity is created analogously, and is calculated separately for each spatial location. For each pollutant its average temporal RSD is larger than its average spatial counterpart, with the former ranging between 43% (NO_2) and 87% (CO), while the latter lies between 32% (NO_2) and 59% (CO). Figure 6-3 shows the distribution of daily spatial RSD for 1997, and the range of values for CO and PM_{10} vary a lot more between days than those for NO_2 and O_3. Figure 6-4 depicts the Greater London boundary, and shows the locations of the monitoring sites (in six figure grid references) for each pollutant. The majority of monitors are located in the middle of the region which is not ideal for a spatial analysis, a point which is taken up in chapter seven. A number of meteorological covariates are also available, and include daily mean temperature, number of sunshine hours and the amount of rain.

Figure 6-3: Distributional estimate of the daily spatial relative standard deviation for each pollutant.

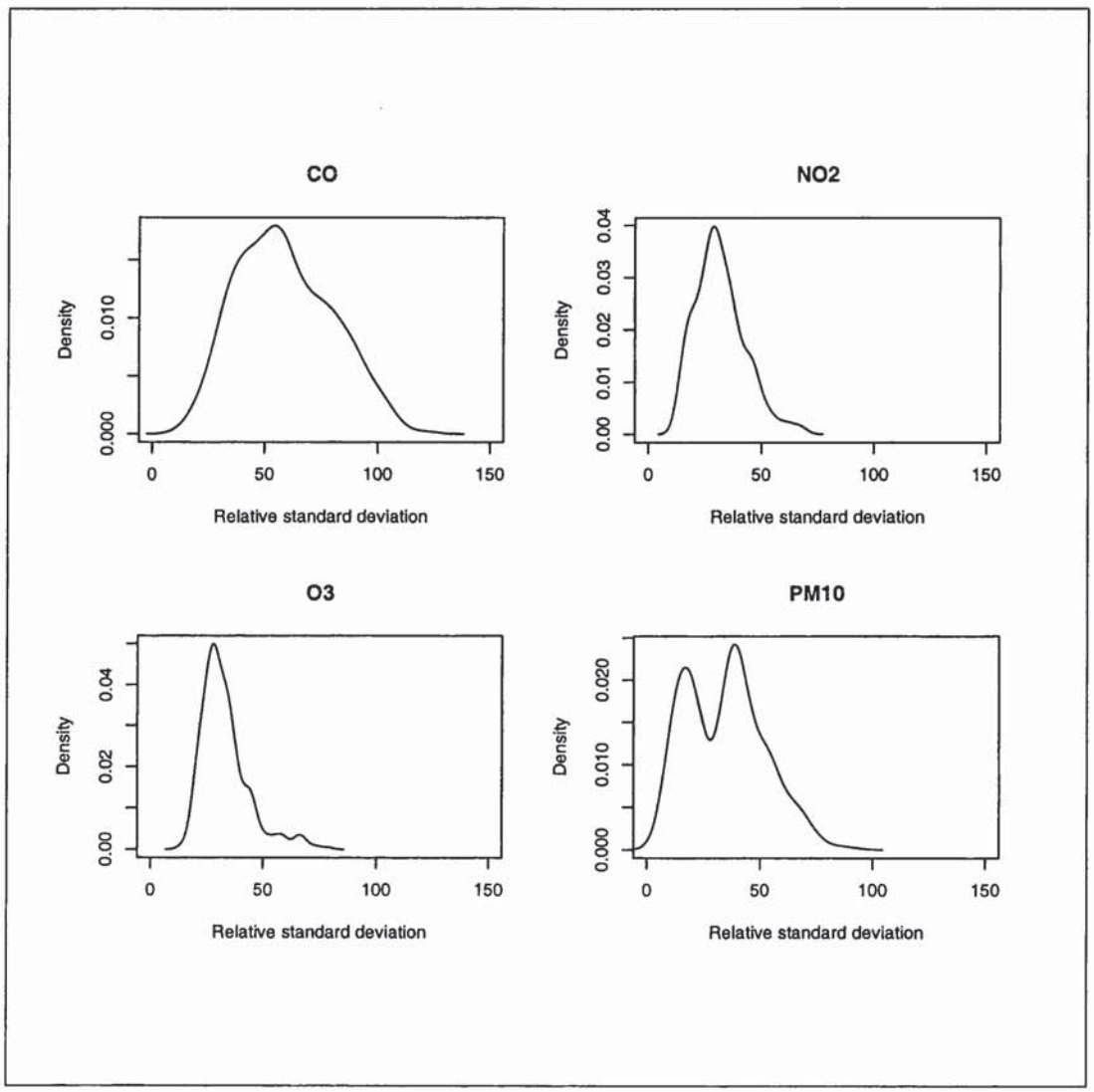


Figure 6-4: Locations of available monitoring sites for each pollutant together with the Greater London boundary.

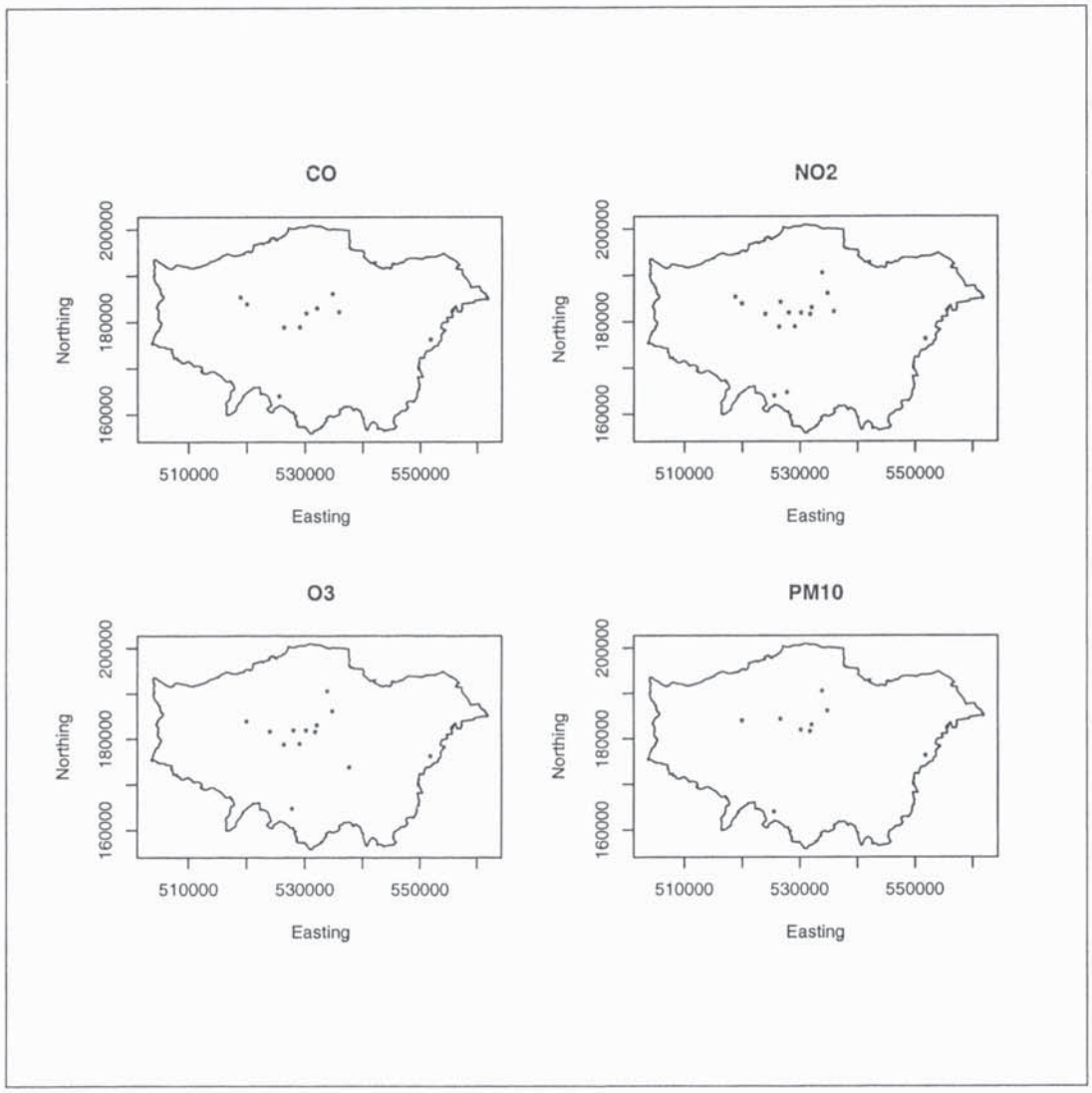


Table 6.5: Summary of the pollution data.

Characteristic	Pollutant			
	CO	NO_2	O_3	PM_{10}
Maximum number of daily monitors	10	16	13	9
Minimum number of daily monitors	7	11	9	6
Mean number of daily monitors	9.5	15.5	12.3	8.6
Percentage missing	4.6	3.3	5.4	3.8
Mean	0.87	27.9	28.1	22.8
Spatial relative standard deviation	59%	32%	33%	35%
Temporal relative standard deviation	87%	43%	63%	57%

6.6.2 Description of the statistical models

The standard, measurement error and spatio-temporal models used in this case study are given below, with the pollutant specific regression component $B_i\delta$ being listed after the model descriptions.

Standard model

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= w_{i-1}\gamma + \alpha_1 + S(i|11, \boldsymbol{\alpha}_2) + S(\text{temperature}_i|3, \boldsymbol{\alpha}_3), \\ \boldsymbol{\beta} = (\gamma, \boldsymbol{\alpha}) &\sim \text{N}(\boldsymbol{\mu}_\beta, \Sigma_\beta). \end{aligned} \tag{6.10}$$

Measurement error model

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\ \ln(\mu_i) &= x_{i-1}\gamma + \alpha_1 + S(i|11, \boldsymbol{\alpha}_2) + S(\text{temperature}_i|3, \boldsymbol{\alpha}_3), \\ \boldsymbol{\beta} = (\gamma, \boldsymbol{\alpha}) &\sim \text{N}(\boldsymbol{\mu}_\beta, \Sigma_\beta), \\ \ln(w_i(\mathrm{s}_l)) &\sim \text{N}(\ln(x_i), \sigma_\epsilon^2) \quad \text{for } l = 1, \ldots, k, \\ \ln(x_i) &\sim \text{N}(\mu_x, \sigma_x^2), \\ \sigma_\epsilon^2 &\sim \text{Inverse-Gamma}(0.01, 0.01). \end{aligned} \tag{6.11}$$

Spatio-temporal measurement error model

$$
\begin{aligned}
y_i &\sim \text{Poisson}(\mu_i) \quad \text{for } i = 1, \ldots, n, \\
\ln(\mu_i) &= x_{i-1}\gamma + \alpha_1 + S(i|11, \alpha_2) + S(\text{temperature}_i|3, \alpha_3), \\
\beta = (\gamma, \alpha) &\sim \text{N}(\mu_\beta, \Sigma_\beta), \\
\mathbf{x} &\sim f(\mathbf{x}|\text{w}(\mathcal{S}), \omega_3), \\
\ln(\mathbf{w}_i(\mathcal{S})) &\sim \text{N}(\ln(\mathbf{x}_i(\mathcal{S})), \sigma_\epsilon^2 \text{I}), \\
\ln(\mathbf{x}_i(\mathcal{S})) &= B_i\delta + \mathbf{1}\theta_i + \mathbf{m}(\mathcal{S}), \\
\theta_i &\sim \text{N}(\phi_1\theta_{i-1}, \sigma_\theta^2) \quad \text{for } i = 2, \ldots, n, \\
\mathbf{m}(\mathcal{S}) &\sim \text{N}(\mathbf{0}, \sigma_m^2\Sigma(\phi_2)), \\
\delta &\sim \text{N}(\mu_\delta, \Sigma_\delta), \\
\sigma_\epsilon^2 &\sim \text{Inverse-Gamma}(0.01, 0.01), \\
\sigma_\theta^2 &\sim \text{Inverse-Gamma}(0.01, 0.01), \\
\sigma_m^2 &\sim \text{Inverse-Gamma}(0.01, 0.01), \\
f(\phi_1) &\propto 1, \\
\phi_2 &\sim \text{Uniform}(0.001013773, 0.05920836).
\end{aligned}
\tag{6.12}
$$

For CO, NO_2 and PM_{10} a lag of one day is used, while O_3 is included at a lag of two days (for reasons see the model building strategy below). For observation $w_i(\mathbf{s}_l)$ the regression component $B_{il}\delta$ for each pollutant is given by

CO $\delta_1 + \delta_2 i + \delta_3 i^2 + \delta_4 \text{E}_l + \delta_5 \text{N}_l + \delta_6 \text{E}_l \times i + \delta_7 \text{N}_l \times i + \delta_8 \text{rain}_i$,

NO$_2$ $\delta_1 + \delta_2 i + \delta_3 i^2 + \delta_4 \text{E}_l + \delta_5 \text{N}_l + \delta_6 \text{N}_l \times \text{E}_l + \delta_7 \text{E}_l \times i + \delta_8 \text{N}_l \times i + \delta_9 \text{rain}_i + \delta_{10} \text{sun}_{i-1}$,

O$_3$ $\delta_1 + \delta_2 i + \delta_3 i^2 + \delta_4 \text{E}_l + \delta_5 \text{N}_l + \delta_6 \text{N}_l \times \text{E}_l + \delta_7 \text{rain}_{i-1} + \delta_8 \text{sun}_{i-1}$,

PM$_{10}$ $\delta_1 + \delta_2 i + \delta_3 i^2 + \delta_4 \text{E}_l + \delta_5 \text{N}_l + \delta_6 \text{N}_l \times \text{E}_l + \delta_7 \text{E}_l \times i + \delta_8 \text{N}_l \times i$.

In the descriptions above, 'rain_i' denotes the amount of rainfall in centimeters, 'sun_i' denotes the hours of sunshine, 'i' denotes calendar time in days, while 'E' and 'N' denote the Easting and Northing six digit grid references of $\mathbf{s}_l$. Further details of the model building process are given below, which includes the motivation for the regression component described above.

Model building strategy

The model building process begins by removing the trend, seasonal variation, over-dispersion and temporal correlation from the respiratory mortality series. In common with the previous chapter these confounding factors are removed with natural cubic splines of calendar

time and temperature, because it is one of the standard approaches adopted in the air pollution literature (see for example Daniels et al. (2004)). The degrees of freedom for the natural cubic splines are chosen by deviance information criteria, which results in eleven for calendar time and three for temperature. The relationship between temperature and mortality was investigated at a number of lags, and the same day's temperature has the lowest DIC (and the most significant effect when included as a linear covariate) and is used in all analyses. The Bayesian residuals for this model appear to be random fluctuations with a fixed variance, suggesting that these covariates are satisfactory. After the covariates were selected, the relationship between each pollutant and mortality was investigated. This relationship was examined at a number of lags using deviance information criteria, and a lag of one day was chosen for CO, NO_2 and PM_{10}, while two days was selected for O_3.

The covariates in the pollution model ($B_i\boldsymbol{\delta}$) were also selected using deviance information criteria, but the model building process began with a single set of five covariates. These covariates were selected by a visual examination of the data, and included an intercept term, linear and quadratic functions of calendar time, and linear trends in Easting and Northing. Each covariate made a noticeable reduction to the DIC for all four pollutants, and are included in each of the pollution models. The effects of space-time interactions and higher order terms were then examined, and easting and northing interactions were added to the NO_2, O_3 and PM_{10} models, while space time interactions (that is easting multiplied by time and northing multiplied by time) were important covariates for CO, NO_2 and PM_{10}. After the interaction terms were included the impact of meteorological covariates was assessed. Daily mean temperature, number of sunshine hours and the amount of rainfall were available, although the first of these is included in the mortality model so cannot be used here. The impact of rainfall and sunshine were investigated at a number of lags for each pollutant, and lags of zero or one day had the lowest DIC in all cases. For CO only the same days rainfall was added to the model, while neither meteorological covariate made a significant impact on PM_{10} levels. Both rain and sunshine had significant impacts on NO_2 and O_3, although the optimum lags were not consistent. In addition the type of monitor (for example, kerbside, in an suburban area, etc) was also available, but its inclusion in any of the models caused non-identifiability problems with the spatial random effects and was not used in the final analysis.

A small amount of each pollution series is missing, which adds an additional complication to the analyses. These values are typically ignored unless a spatio-temporal pollution model is used, which effectively assumes these data are missing at random. This is the approach taken here for models (6.10) and (6.11), because accurate prediction of these values would be nearly impossible. However for the spatio-temporal model the missing observations can be predicted with greater accuracy, because they can be based on the spatial location of an observation as well as the day it relates to. Consequently for model (6.12)

the true values ($x_i(\mathbf{s}_j)$) of the missing observations are predicted as part of the MCMC algorithm, and are re-estimated at each iteration. The spatially averaged pollution levels $\mathbf{x}$ are estimated at each iteration of the MCMC algorithm as described in section 6.4.1, by predicting the true pollution surface at 90 prediction locations. These locations make up a square grid of points 4km apart (4,000 metres on the Easting and Northing scale), and the corresponding predictions are averaged to obtain $\mathbf{x}$.

For the measurement error and spatio-temporal models, sensitivity analyses were conducted to ascertain whether the choice of priors affected the results. These analyses focused on prior specification for the variance and correlation parameters, because the other parameters are given standard priors that should work well in a variety of situations. A flat prior on the scale of standard deviation (effectively an inverse-gamma$(-0.5, 0)$ distribution) and a series of inverse-gamma(ϵ, ϵ) priors (where $\epsilon = 0.1, 0.001$) were specified for each variance parameter, but the results were robust to these changes. Consequently an inverse-gamma$(0.01, 0.01)$ prior was used in all cases, because unlike a flat prior on the standard deviation scale (suggested by Gelman (2006)) it is a proper distribution. A uniform prior is specified for the spatial correlation parameter ϕ_2, because nothing was known about its value before the study was undertaken. The limits on the prior are fixed so that the minimum correlation between any two points in Greater London is between 0.05 and 0.95, allowing the pollution surface to exhibit a large range of correlations. The largest distance between two locations in Greater London is 50.6km (50,600 metres), which leads to $p_1 = 0.001013773$ and $p_2 = 0.05920836$. Numerous alternative limits were investigated, but the resulting relative risks showed minimal changes. In common with the simulation study the correlation parameter did not converge, and its credible interval was close to the prior limits. The possible reasons for this are discussed in chapter seven.

6.6.3 Results

The models contain a large number of parameters, so to aid convergence the mortality and pollution covariates are standardised to have a mean of zero and a standard deviation of one before inclusion in the model (and are subsequently back-transformed when obtaining results from the posterior distribution). Inference is based on a single Markov chain, which is burnt in for 40,000 iterations at which point convergence was assessed to have been reached using the methods of Gelman et al. (2003). The chain is then run for an additional 100,000 iterations which are thinned by five to reduce the correlation, resulting in 20,000 samples from the joint posterior distribution. In this section, let the subscripts 'ST', 'ME' and 'SP' denote estimates from the standard, measurement error and spatio-temporal models respectively.

Table 6.6: Correlations (upper diagonal) and average differences (lower diagonal) in estimated exposures between each of the three models

		Standard	Measurement error	Spatio-temporal
CO	Standard		0.9954	0.9939
	Measurement error	0.11		0.9941
	Spatio-temporal	0.11	0.21	
NO_2	Standard		0.9964	0.9947
	Measurement error	1.38		0.9982
	Spatio-temporal	4.94	6.25	
O_3	Standard		0.9988	0.9987
	Measurement error	1.42		0.9993
	Spatio-temporal	2.60	3.98	
PM_{10}	Standard		0.9550	0.9616
	Measurement error	2.87		0.9697
	Spatio-temporal	3.91	2.29	

(i) - Pollution estimates

The pollution estimates are summarised in Figure 6-5, Figure 6-6 and Table 6.6, which are based on pairwise comparisons because the true pollution levels are unknown. The Figures depict pairwise scatter plots of the estimates, while the table presents their correlations and mean absolute differences. For CO, NO_2 and O_3 the estimates exhibit clear differences in size, with $\hat{x}_i^{ME} < \hat{x}_i^{ST} < \hat{x}_i^{SP}$ for at least 82% of days. In contrast for PM_{10} the standard model produces the largest estimates, while estimates from the measurement error and spatio-temporal models are of a similar size (56% of days had a larger measurement error estimate). These results are re-enforced by Table 6.6, which shows that the biggest mean absolute difference is between the Measurement error and spatio-temporal estimates (not true for PM_{10}). For each pollutant the differences between $\hat{x}_i^{ST}$ and $\hat{x}_i^{ME}$ are fairly constant, and do not depend on the estimated pollution level. However this is not true for the other pairwise comparisons (that is ST against SP and ME against SP), where the differences increase for higher pollution levels. This indicates that all three estimates are similar if the pollution level is low, while the values from the spatio-temporal model become less similar to the other two as the pollution level increases. Table 6.6 shows the correlations between each pair of estimates, which is close to one in all cases. However the correlations are significantly lower for PM_{10}, which may be caused by the increased amount of measurement error in these data (see Table 6.7). Table 6.7 also shows that the CO data have the largest combined amount of measurement error and spatial variation, which re-enforces the differences in spatial RSD from Table 6.5. The remaining pollutants have a similar percentage of spatial RSD, which is re-enforced by similarities in their combined amounts of spatial variation and measurement error (see Table 6.7).

Figure 6-5: Pairwise scatter plots of all three estimates of **x**. The left column relates to CO while the right column relates to NO_2. The three estimates are denoted by: ST - standard, ME - measurement error and SP - spatio-temporal respectively.

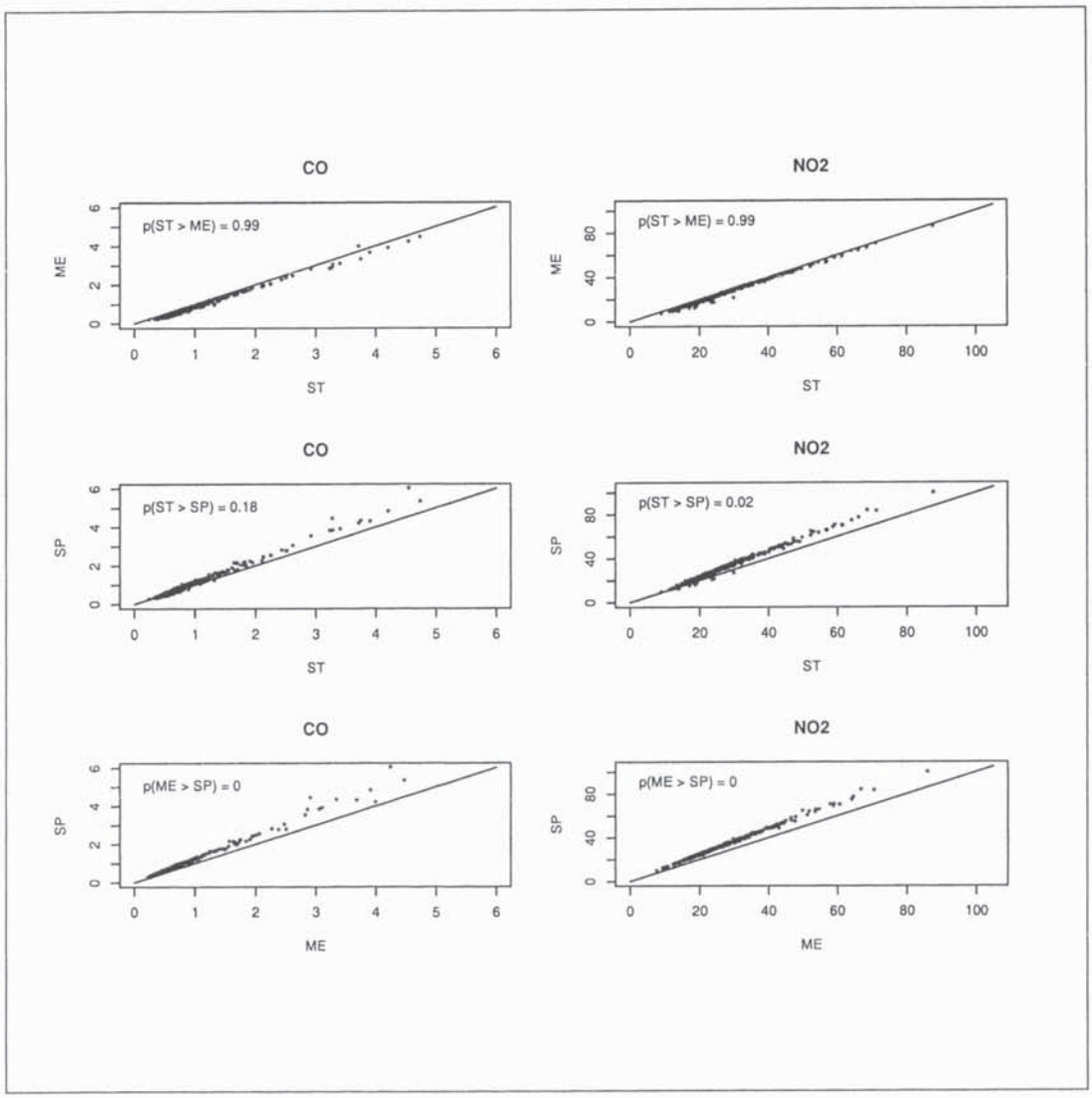


Table 6.7: Estimated levels of measurement error and spatial variation on the natural log scale with 95% credible intervals.

Model	pollutant			
	CO	NO_2	O_3	PM_{10}
Measurement error σ^2_ϵ	0.387 (0.381, 0.407)	0.163 (0.157, 0.169)	0.168 (0.161, 0.175)	0.477 (0.453, 0.503)
Spatio-temporal σ^2_ϵ	0.143 (0.136, 0.150)	0.090 (0.087, 0.093)	0.100 (0.096, 0.104)	0.436 (0.415, 0.460)
σ^2_m	1.071 (0.422, 1.936)	0.848 (0.382, 1.860)	0.504 (0.200, 1.702)	0.134 (0.040, 1.204)

Figure 6-6: Pairwise scatter plots of all three estimates of $\mathbf{x}$. The left column relates to O_3 while the right column relates to PM_{10}. The three estimates are denoted by: ST - standard, ME - measurement error and SP - spatio-temporal respectively.

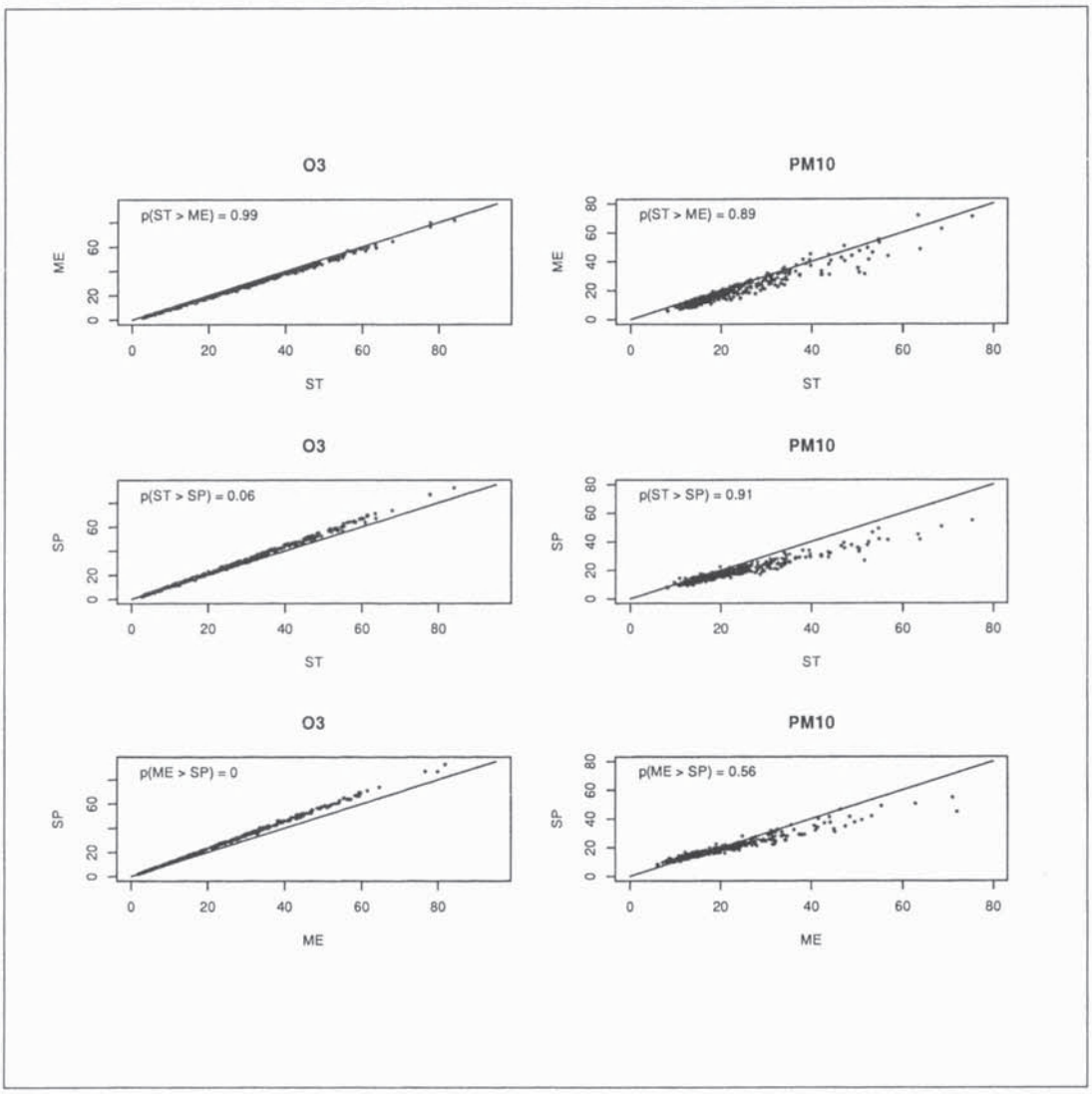


Table 6.8: Estimated relative risks and 95% credible intervals for each data set and model. The relative risks relate to an increase of ten units of pollution for NO_2, O_3, PM_{10} and one unit for CO.

Pollutant	Model		
	Standard	Measurement error	Spatio-temporal
CO	1.0283 (1.0006, 1.0566)	1.0294 (0.9997, 1.0603)	1.0186 (0.9980, 1.0494)
NO_2	1.0197 (1.0019, 1.0371)	1.0210 (1.0022, 1.0373)	1.0158 (1.0016, 1.0352)
O_3	1.0177 (1.0033, 1.0323)	1.0185 (1.0032, 1.0353)	1.0150 (1.0025, 1.0306)
PM_{10}	1.0250 (1.0080, 1.0421)	1.0365 (1.0187, 1.0586)	1.0326 (1.0083, 1.0601)

(ii) - Results of the pollution - mortality associations

The estimated relative risks for each pollutant and model are presented in Table 6.8, and are of a similar size to those found in previous studies (see for example Dockery and Pope (1994) and Samet et al. (2000)). For NO_2, O_3 and PM_{10} the risks relate to an increase of ten units of pollution, while for CO the increase is for one unit because the underlying pollution level is much lower (0.87). The maximum difference between each set of three estimates is: CO 1%, NO_2 0.5%, O_3 0.35%, PM_{10} 1.5%, which for CO and PM_{10} are an increase in risk of one and a half times. For all four pollutants the relative sizes of the three risks are opposite to the ordering of the estimated pollution levels, meaning that $x_i\gamma$ is roughly constant across all three models. This suggests that as exposure increases the relative risk falls, meaning that correctly estimating exposure is a crucial part of any analysis. It also implies that the differences between (RR_{SP}, RR_{ST}, RR_{ME}) will depend on the differences in ($\hat{x}_i^{ME}, \hat{x}_i^{ST}, \hat{x}_i^{SP}$), although the sizes cannot be compared across pollutants because each has a different underlying mean level.

6.7 Discussion

This chapter investigates how mis-estimating pollution exposures can affect the relative risk, with particular interest in how the standard method compares with more complex alternatives. The simulation study shows that the standard estimate performs relatively well in all scenarios, with maximum biases of less than 1% on the relative risk scale. For pollution data that are spatially flat and exhibit no measurement error this simple average is always accurate, causing little or no bias in the estimated relative risk. In all other situations the standard estimate is much less accurate, although its performance for a single data set is unpredictable. This unpredictability is probably due to the reaction between the locations of the monitoring sites and the peaks and troughs in the underlying pollution surface, meaning that the standard estimate may produce accurate results even in

the presence of spatial variation and measurement error. However on average both factors reduce its accuracy significantly, with the bias in relative risk increasing by nine (spatial variation) and eighteen (measurement error) times respectively. In addition a high level of measurement error typically biases the relative risks towards one (no association with mortality), whereas spatial variation has no such directional effect. Furthermore, adding measurement error to spatially variable pollution data doubles the bias in relative risk, whereas the addition of spatial variation to data contaminated by measurement error has little or no impact. These findings indicate that measurement error is more detrimental to the accuracy of the standard estimate than spatial variation, which is not surprising given that the true exposure is the average pollution level across the region of interest. If a pollution surface is spatially variable but contains no measurement error, then as long as the monitoring sites are not all located at peaks (or troughs) in the pollution surface, averaging these values should produce estimates that are close to the true value. In contrast if a pollution surface is spatially flat but contaminated by measurement error, the standard estimate is an average of a small number (less than 20) of noisy observations, which may not be close to the true value. The simulation study also showed that if spatial variation and measurement error are not present, then the number of monitors has little affect on the accuracy of the standard estimate. However, if either factor is present the average bias halves if twenty spatial observations are available compared with only five.

The simulation study shows that for flat pollution surfaces with no measurement error the choice of model is immaterial, as the standard estimate produces similar results to the measurement error and spatio-temporal models. However, if either factor is present the spatio-temporal model nearly always outperforms the standard approach, yielding estimates of exposure and relative risk that are less biased. The measurement error model typically performs better than the standard approach but less well than the spatio-temporal model, and is only worth considering if the pollution data have a high degree of measurement error and minimal spatial variation. However in this situation all three models perform poorly, and the differences between them are not large. In contrast, if the pollution data are dominated by spatial variation the differences between each model are more pronounced, and the spatio-temporal model nearly always produces less biased results. These improvements are still observed if only five pollution monitors are available, implying that five sites is enough for a spatio-temporal model to produce better results than the standard average. However the number of monitors has a major impact on the accuracy of all three models, with biases in exposure and relative risk dropping by more than half if twenty monitors are available compared with only five.

The differences between the three approaches are more pronounced in the case study than in the simulation study, as the former reports relative risks differing by between 0.35% and 1.5%, depending on the pollutant analysed. This is probably due to the increase in

spatial variation and measurement error in these real data, which is estimated to be between 0.09 and 1.07 (by the spatio-temporal model) compared with a maximum of 0.0289 for the simulation study. The true exposure and relative risk are unknown for these real data, and evidence from the simulation study indicates that the spatio-temporal estimates are likely to be the most accurate. If this is true the standard approach may be mis-estimating the relative risk by around 0.97% (CO), 0.39% (NO_2), 0.27% (O_3) and -0.76% (PM_{10}) for these data, which is a significant amount considering the average risk in air pollution studies is between 1% and 2%. Furthermore, these results imply that authors who use the standard estimate may be reporting relative risks that are biased by up to 1%. However, the risks from the spatio-temporal model are both higher (PM_{10}) and lower (CO, NO_2, O_3) than those obtained using the standard approach, meaning that the true associations are unlikely to be over or under reported in general. These differences in relative risk are probably caused by variation in the estimated exposures, because $x_i\gamma$ remains roughly constant for each of the three models. This implies that mis-estimating exposure causes the relative risk to be mis-estimated in the other direction, meaning that correct estimation of pollution levels is a vital component of any study. As expected the exposures estimated from each model are very similar, with the only significant differences occurring on days with high pollution levels. On these days the spatio-temporal model produces estimates that are both larger (PM_{10}) and smaller (CO, NO_2, O_3) than the other two approaches, which is most likely caused by the relationship between the locations of the monitoring sites and the peaks and troughs in the underlying pollution surface.

This chapter has shown that correctly estimating exposure is a vital part of any study, and that a poor estimate can cause the relative risk to be biased by a significant amount. Three alternative estimates have been compared using real and simulated data, and the choice between them requires a trade-off of simplicity against accuracy. The differences in accuracy depend on the pollution data, and in particular its amount of spatial variation and measurement error. If both factors are low then each model is adequate, and the standard estimate can be used with its advantage of simplicity. In contrast, as both factors increase the bias when using the standard method will rise, and in such cases analysis should be based on the spatio-temporal model. The combined amounts of spatial variation and measurement error can be measured by the spatial relative standard deviation, and for data sets with values around one percent (the value for the simulation study data in scenarios one to four) the standard approach should produce adequate results. In contrast for data sets with values above fifteen percent (scenarios five to sixteen from the simulation study gave values around fifteen percent, while the real data in the case study have much higher values), the standard approach is likely to produce significantly biased results, and a spatio-temporal model should be used.

Chapter 7

Conclusion

This thesis focuses on the relationship between short-term exposure to air pollution and mortality, which has been a public health concern for over fifty years. The majority of studies investigating this relationship are based on ecological data, and estimate a group level association between ambient pollution levels and population aggregated mortality rather than an exposure-response relationship. These data are the only type that are widely available, but present researchers with numerous statistical challenges. These challenges are outlined in chapters two and three, which review the statistical methodology used in air pollution and health studies as well as the related literature. This review also critiques the standard approaches to estimating the pollution-health association, which are then extended by the methodology developed in chapters four to six. These chapters examine substantive issues in modelling air pollution and health data, and in each case a modelling framework is proposed that allows specific questions to be addressed about the underlying processes that generate these data. The efficacy of the developments proposed here are assessed using a variety of real and simulated data, and the motivation for each chapter is briefly reviewed below.

Chapter four models air pollution and mortality data with a Bayesian dynamic generalised linear model, which extends the standard approaches to modelling in two ways. Firstly any long-term trends, seasonal variation, over-dispersion or temporal correlation present in the mortality data can be removed with an autoregressive process, which contrasts with the standard approach of using a smooth function of calendar time. Secondly the pollution-mortality relationship can also be represented as an autoregressive process, allowing it to change over time rather than being fixed at a constant value. Chapter five proposes a time-varying coefficient model for air pollution and health data, which examines whether the pollution-mortality relationship changes over time. This model extends the investigation presented in chapter four, by modelling the temporal evolution as a smooth function. This representation is more parsimonious than the autoregressive process adopted in chapter four, reducing the problems of non-identifiability which

lessens the need for highly informative prior distributions. Chapter six examines how mis-estimating air pollution exposure can affect the resulting relative risk, and assesses the performance of the standard approach that averages the spatial pollution observations in relation to more complex alternatives. These alternatives are based on measurement error and spatio-temporal models, which incorporate the factors likely to affect the accuracy of the standard approach. The developments proposed in this thesis focus on three main themes, which incorporate both the air pollution and covariate components of a statistical model. The first of these focuses on how unmeasured risk factors that affect the mortality data are modelled, and specifically whether an autoregressive process or a smooth function is more appropriate. The second theme relates to the pollution-mortality relationship, and in particular whether it changes over time. The third is concerned with pollution exposure, and how mis-estimating it affects the resulting health risk. This work has led to a number of other recurring themes, which are a byproduct of the analyses rather than a subject of *a-priori* interest. The first of these is how Bayesian and likelihood approaches to analysis compare in applied problems, while the second is the impact of prior distributions for variance parameters. Each of these themes is discussed in detail below, beginning with how unmeasured risk factors should be modelled.

7.1 Key theme - Modelling the influence of unmeasured risk factors

The mortality data analysed in this thesis exhibit long-term trends, seasonal variation, over-dispersion and temporal correlation, which are induced by covariate risk factors that affect the number of daily deaths. A proportion of these factors are meteorological such as temperature or dew-point temperature, whose effects can be removed by including them as covariates in the regression model. The remaining factors are largely unknown, and their influence is typically removed by adding pairs of trigonometric terms at a number of frequencies or a smooth function of calendar time to the regression model. The majority of recent studies have used smooth functions of calendar time for modelling such factors, which are typically represented as parametric regression splines (Daniels et al. (2004)) or non-parametric smoothing splines (Dominici et al. (2000)). In chapter four the efficacy of dynamic models are investigated in this setting, which allows autoregressive processes to be used in place of the smooth functions. This extension to the standard model is initially appealing, because autoregressive processes share a discrete time support with the daily mortality counts, only having a single value for each day. In contrast smooth functions have a continuous time support, meaning that they have a value for any point in time. The efficacy of using autoregressive processes to model the underlying trend in daily mortality has been investigated here, focusing on random walks of first and second order as well as a local linear trend model.

The natural cubic spline and three autoregressive processes capture the underlying trend in daily mortality well, and the choice between them has little impact on the estimated health risk associated with pollution exposure. The only differences between these models are how well they represent the yearly mortality peaks, and the amount of temporal correlation they leave in the residual series. The second order random walk proved to be the best of the four alternatives, modelling the peaks in mortality closely and leaving no correlation in the residuals. The next best was the local linear trend model, which only exhibited significant correlation at the first lag. The natural cubic spline and first order random walk gave similar results, with the former leaving significant correlation in the residual series. The second order random walk and local linear trend model are the most flexible of the four models used here, which is most likely the reason for their superior performance. However there is a possible source of bias in the Bayesian analyses, as the autoregressive smoothing parameters are estimated as part of the MCMC algorithm, whereas the degrees of freedom for the spline is fixed prior to estimation. This allows the autoregressive trends to incorporate variation in their respective smoothing parameters, which may be the reason for their superior performance compared with the spline. To solve this problem a penalised approach to estimation could be adopted for the natural cubic spline, which would allow the smoothing parameter to be estimated within the MCMC algorithm (as in chapter five). However there are no such differences in the likelihood based analyses, where all four smoothing parameters are estimated using Akaike's information criteria. In the analysis presented here the natural cubic spline still exhibits the worst performance, indicating that a second order random walk should be preferred in future studies. However these results only refer to one data set, and further research is required to determine whether these conclusions hold more generally. The poor performance of the natural cubic spline may also be caused by its lack of flexibility, and a comparison between the autoregressive processes adopted here and a more general class of non-parametric smooth functions would be an interesting research topic. The increased flexibility of such functions compared with natural cubic splines are likely to result in an improved performance in this regard, although whether it would outperform a second order random walk is unclear.

The analysis presented in chapter four compares both Bayesian and likelihood approaches to estimation, which is of interest because the only previous use of dynamic models in an air pollution context adopts maximum likelihood estimation. The results show that the Bayesian estimates outperformed their likelihood counterparts for all four trend models, capturing the winter peaks in mortality better and leaving significantly less correlation in the residual series. This indicates that for the three autoregressive processes the Kalman filter over-smoothes the mortality series, where as the autoregressive prior results in a more flexible trend with less residual correlation. Although the Bayesian and likelihood

estimates exhibit different performances the autoregressive processes outperform the natural cubic spline in both settings, indicating that the use of dynamic modelling in this context has great potential and should be considered for future studies.

7.2 Key theme - Temporal variation in the pollution-health relationship

Chapters four and five investigate whether the pollution-mortality relationship exhibits any temporal variation, by analysing data from Greater London, Cleveland, Detroit, Minneapolis and Pittsburgh. The relationship in the first of these cities is represented as a first order random walk within a dynamic generalised linear model, while the remainder are estimated in chapter five using a variety of functions. These functions include a penalised natural cubic spline, a first order random walk, a cubic polynomial, the cyclical seasonal form proposed by Peng et al. (2005) and a constant function, all of which estimate relative risks of approximately 0.5%. The temporal variation in each of these relationships is small, and a constant association is most likely in the majority of these cities. However in Pittsburgh the relative risk has a range of 2%, which the simulation study has shown is unlikely to be a mis-estimated constant association. Consequently the possibility of temporal variation in the pollution-mortality relationship deserves further consideration, and one avenue would be to use longer time series to see whether such a trend continues over a longer period, for example fifteen to twenty years.

The cubic and seasonal models are not appropriate for this problem, because they are rigid parametric forms and cannot incorporate temporal variation with different shapes. In chapter five the cubic model appears to perform well while the seasonal form gives spurious results, which is because the true relationships are likely to exhibit slowly evolving trends. In contrast the first order random walk model does not have a pre-determined shape, which makes it an improvement on these simple parametric forms. However it represents the relationship with an overly large number of parameters, causing non-identifiability problems that result in estimates being contaminated with unwanted noise. This non-identifiability can be removed by assigning the random walk variance a highly informative prior, which reduces the number of effective parameters by constraining consecutive elements to have similar values. This means that the smoothness of the estimated relationship is highly dependent on the informativeness of the prior, allowing a researcher to choose the desired amount of variability to fit their required hypothesis. This prior dependence is clearly undesirable, and a more parsimonious representation would be preferable. This is provided by the penalised spline model proposed in chapter five, which combines the smoothness of the simple parametric forms with the flexibility of the random walk.

The analyses of chapters four and five have compared the efficacy of smooth functions and autoregressive processes for two distinct problems: (i) modelling the trend and correlation in the mortality data; (ii) representing the temporal variation in the pollution-mortality relationship. These analyses indicate that a random walk outperforms a smooth function for the first of these problems, while the opposite is true for the second. The preference for an autoregressive process or smooth function depends on the amount of variation that needs to be modelled, which is greater for problem (i) than it is for problem (ii). A random walk is typically more variable than a smooth function because it has a larger number of free parameters, meaning that is more adept at modelling large amounts of non-linearity but less effective when the variation is small.

The random walk and penalised spline models adopted in chapters four and five are applied in both Bayesian and likelihood settings, with the former producing estimates that exhibit greater curvature in nearly all cases. There are a number of possible reasons for this increased variation estimated by the Bayesian approach, one of which is how their respective smoothing parameters are estimated. As previously described the likelihood approach to estimating smoothing parameters fixes them at the chosen value, while the Bayesian approach simultaneously estimates them within the MCMC scheme. For the autoregressive processes there are additional differences between the Bayesian and likelihood approaches to estimation, as the latter applies additional smoothing using the Kalman filter. Either of these factors may be causing the difference in curvature between the Bayesian and likelihood estimates, although as the latter explanation only applies to dynamic models, the former seems more likely.

The time-varying relationships presented in chapters four and five relate daily mortality to pollution exposure, the latter of which was estimated using the standard method. However the conclusions from chapter six suggest that this simple average may be mis-estimating exposure if the pollution surface exhibits spatial variation or measurement error, and it would be interesting to investigate whether the time-varying associations observed here retain their shape if an alternative measure of pollution exposure is used. In particular a future research avenue would be to combine the time-varying pollution-mortality association with a spatio-temporal pollution model, to determine if the same long-term trends over time are observed. However such extensions may not be feasible, as both the time-varying relationships and the spatio-temporal pollution model are computationally demanding. Another interesting area of future research would be to apply the time-varying coefficient model to data from multiple cities simultaneously, allowing regional and national time-varying relationships to be estimated. Other research avenues include modelling the time-varying associations with more flexible smooth functions such as smoothing splines, though as the time-varying relationships in this thesis evolve relatively slowly this seems unlikely to produce improved results. Furthermore the pollution-mortality relationship could be allowed

to depend on other covariates such as temperature, though as the relationships estimated here did not exhibit seasonal behaviour this possibility seems unlikely.

7.3 Key theme - Estimating air pollution exposure

Chapter six investigates how mis-estimating pollution exposure impacts on the relative risks, comparing the standard average with measurement error and spatio-temporal models. The simulation and case studies indicate that correctly estimating pollution exposure is a crucial part of any study, with poor estimates likely to result in biased health risks. This bias appears to be in the opposite direction to that of pollution exposure, indicating that the pollution component of a regression model, $x_i\gamma$, remains roughly constant. The standard estimate of pollution exposure performs well in the majority of situations, but gets less accurate as the amount of measurement error and spatial variation increases as might be expected. The impact of measurement error is more pronounced than that of spatial variation, causing the average bias to double while shrinking the relative risks towards one on average. All three models perform well if measurement error and spatial variation are low, and the standard average is preferable because of its simplicity. However, if either factor becomes non-negligible the spatio-temporal model typically outperforms the other two, and is likely to produce relative risks that are less biased. For the data analysed the improvement may be as large as 1% on the relative risk scale, which is a significant difference considering the majority of studies estimate risks that are less than 2%. At present only a small number of studies use spatio-temporal pollution models for this problem, but the evidence presented here suggests that such models should be routinely adopted. However estimating exposure as the daily average of pollution measurements does not lead to over or under estimated health risks in general, indicating that the majority of current studies should be reporting associations that are likely to be close to the correct size.

As previously mentioned the majority of researchers estimate pollution exposure by averaging the spatial observations, which may be because they do not have access to the raw data or due to the simplicity of this approach. However the deficiencies of this simple method should be investigated further, with particular interest in how it compares to more complex alternatives. The simulation study presented in chapter six only considers two levels of spatial variation and measurement error, and a more in-depth extension seems an exciting area of future research. Such work should estimate the relationships between levels of spatial variation and measurement error against average bias in relative risk, with particular interest in whether the resulting curves are smooth or contain sudden changes. The spatio-temporal model adopted here specifies an exponential correlation structure for the spatial observations, and more complex non-separable, non-stationary and non-isotropic models are a natural avenue of future work. However the correlation model used in chapter six had convergence problems, which is likely to be exacerbated

if more complex alternatives are used. The non-convergence observed here is most likely caused by non-identifiability of ϕ_2, whose full conditional distribution is relatively flat. The cause of this non-convergence needs to be addressed, and another avenue of future work is to determine if this problem persists as the number of spatial observations increase.

In chapter six the bias in relative risks from all three models halved if twenty monitors were used compared with only five, and it is of interest to conduct a more thorough examination to estimate the relationship between the number of monitoring sites and bias in health risk. Such a relationship would largely depend on the amount of spatial variation and measurement error exhibited by the observed pollution data, and it seems likely that it would be relatively flat if neither factor is present. In contrast if either factor is large the bias in health risk is likely to decrease as the number of pollution monitors increases, with steeper curves observed for larger amounts of measurement error or spatial variation. The interaction between this relationship and the locations of the monitoring sites also warrants attention, because they are rarely equally spaced throughout the region under study, a fact that is likely to hamper accurate pollution estimation. Instead their locations are chosen for strategic reasons, such as cost, availability of space, or at known locations of high (or low) pollution levels. Therefore determining an optimal monitor placement scheme to achieve the least bias in relative risks would be of regulatory interest.

7.4 Related theme - Comparison between Bayesian and likelihood approaches to analysis

The analyses presented in this thesis are predominantly Bayesian, because it is the natural framework in which to view hierarchical data and models. This is because it incorporates variation at multiple levels in a straightforward manner, while basing estimation on simpler sub-models using MCMC simulation. However in chapters four and five likelihood based implementations of most models are also presented, which allows a direct comparison with existing air pollution and mortality studies. These comparisons highlight two major differences between Bayesian and likelihood estimation in applied problems, which are consistently observed in different contexts. The first of these is the amount of curvature exhibited by corresponding estimates of the same function, which is always larger in the Bayesian analyses. This increase in curvature is observed for the time-varying pollution-mortality relationships in chapters four and five as well as the trend models in chapter four, and is most likely caused by the estimation of the respective smoothing parameters as mentioned above. In the likelihood analysis the estimated trend or time-varying relationship is estimated for numerous values of the smoothing parameter, which is then selected by optimising a model selection criterion. In contrast the Bayesian estimates are based on the entire posterior distribution of the smoothing parameter, which includes values that

correspond to large amounts of variation causing the actual estimates to be less smooth. However these differences are also observed for the natural cubic spline trends in chapter four, whose smoothing parameters are fixed prior to analysis by AIC and DIC respectively, meaning that this phenomenon requires additional investigation.

The second difference between the two approaches is the width of their uncertainty intervals, with the Bayesian credible intervals being wider than their likelihood counterparts. This is consistently observed for all the pollution mortality associations, and may also be due to the estimation of the respective smoothing parameters. Consequently the Bayesian intervals are more realistic estimates of the true variability as they incorporate the uncertainty in the smoothing parameters, while their likelihood counterparts are likely to be too narrow. However this phenomenon is also observed for the constant pollution-mortality associations presented in chapter four, even though there are no smoothing parameters in this case. This indicates that posterior simulation may produce wider intervals than those based on a likelihood information matrix.

7.5 Related theme - Prior specification for variance parameters

Variance parameters are an important part of the models proposed in chapters four to six, acting mainly as smoothing parameters for random walks, penalised splines or spatial random effects. In the statistical literature these parameters are typically assigned conjugate inverse-gamma(ϵ, ϵ) priors, which are thought to be non-informative for small values of ϵ such as 0.01. However chapter five highlights that such priors are hardly non-informative for very small values, having little or no mass below 10^{-6}. Consequently the choice of ϵ has a large impact on the time-varying associations represented with a penalised spline, becoming smoother as $\epsilon \rightarrow 0$. This is because as $\epsilon \rightarrow 0$ an inverse-gamma prior has a larger proportion of mass at these very small values, which does not force the variance parameter away from zero. The analyses presented here adopt the proposal of Gelman (2006) and use a flat prior on the scale of standard deviation (effectively an inverse-gamma(-0.5,0) distribution), because it does not force the variance parameter away from zero. In contrast the spatial random effects in chapter six have much larger variances (around 10^{-2}), and the choice of inverse-gamma(ϵ, ϵ) prior has no impact on the results. This indicates that for variance parameters with little or no posterior mass below 10^{-2} the choice of non-informative prior is not important, but for smaller values a flat prior on the scale of standard deviation may be preferred.

The priors described above are all non-informative, but a highly informative prior with a large proportion of mass near zero was required to smooth the random walk processes.

Selecting such a prior from the inverse-gamma family is not straightforward, because it has two parameters that affect its range of likely values. In this thesis I use a zero-mean Gaussian prior that is truncated to be positive, whose informativeness is only controlled by its variance. As this gets smaller the prior mass is forced closer to zero, which makes it more informative and results in smoother trends. This prior worked well for the random walks adopted in chapters four and five, and should be the prior of choice in such circumstances.

7.6 Summary

In summary, the models developed and the conclusions drawn in this thesis should lead to more accurate estimates of the relationship between air pollution exposure and mortality or morbidity. Furthermore it should also lead to a greater understanding of the underlying processes that generate these data, as well as the effects that unrealistic assumptions may have on both the statistical modelling process in general, and the epidemiological enquiry being undertaken. However statistical analyses of the type presented here are faced with a paradox relating to the available data and the choice of statistical model. As highlighted in chapter six, the quantity of data required to produce reliable results from complex models can be vast, typically much more than are available in practice. In contrast if such quantities of data are available the computation is prohibitively large, meaning that reliable estimates are still difficult to obtain. As computing power increases researchers will be able to fit more realistic models without having to make as many (possibly) untenable assumptions about the underlying processes and thus produce more reliable estimates of the relationships between pollution and health.

Bibliography

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory 1*, 267–281.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica 33*, 178–196.

Ameen, J. and P. Harrison (1985). Normal Discount Bayesian Models. *Bayesian Statistics 2*, 271–198.

Banerjee, J., B. Carlin, and A. Gelfand (2003). *Hierarchical Modelling and Analysis for Spatial Data* (1st ed.). Chapman and Hall/CRC, London.

Bates, D. and D. Watts (1988). *Nonlinear Regression Analysis and its Applications* (1st ed.). John Wiley and Sons, Winchester.

Berry, S., R. Carroll, and D. Ruppert (2002). Bayesian Smoothing and Regression Splines for Measurement Error Problems. *Journal of the American Statistical Association 97*, 160–169.

Booth, J., G. Casella, H. Friedl, and J. Hobert (2003). Negative Binomial Loglinear Mixed Models. *Statistical Modelling 3*, 179–191.

Box, G. (1980). Sampling and Bayes' Inference in Scientific Modelling and Robustness. *Journal of the Royal Statistical Society Series A 143*, 383–430.

Box, G. and G. Jenkins (1976). *Time series analysis forecasting and control* (Revised ed.). Holden-Day, San Francisco.

Breslow, N. and D. Clayton (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association 88*, 9–25.

Brumback, B., L. Ryan, J. Schwartz, L. Neas, and P. Stark (2000). Transitional Regression Models, with Application to Environmental Time Series. *Journal of the American Statistical Association 95*, 16–27.

Buckeridge, D., R. Glazier, B. Harvey, M. Escobar, C. Amrhein, and J. Frank (2002). Effect of Motor Vehicle Emissions on Respiratory Health in an Urban Area. *Environmental Health Perspectives 110*, 293–300.

Burke, J., M. Zufall, and H. Ozkaynak (2001). A Population Exposure Model for Particulate Matter: Case Study Results for $PM_{2.5}$ in Philadelphia, PA. *Journal of Exposure Analysis and Environmental Epidemiology 11*, 470–489.

Burnett, R., R. Dales, M. Raizenne, D. Krewski, P. Summers, G. Roberts, M. Raad-Young, T. Dann, and J. Brook (1994). Effects of Ambient Levels of Ozone and Sulfates on the Frequency of Respiratory Admissions to Ontario Hospitals. *Environmental Research 65*, 172–194.

Cakmak, S., R. Burnett, and D. Krewski (1999). Methods for Detecting and Estimating Population Threshold Concentrations for Air Pollution-Related Mortality with Exposure Measurement Error. *Risk Analysis 3*, 487–496.

Carlin, B., H. Xia, O. Devine, P. Tolbert, and J. Mulholland (1999). Spatio-Temporal Hierarchical Models for Analyzing Atlanta Pediatric Asthma ER Visit Rates. *Case Studies in Bayesian Statistics 4*, 303–320.

Carroll, R., D. Ruppert, and L. Stefanski (2006). *Measurement Error in Nonlinear Models* (2nd ed.). Chapman and Hall/CRC, London.

Chatfield, C. (1996). *The Analysis of Time Series: An Introduction* (5th ed.). Chapman and Hall/CRC, London.

Chiogna, M. and C. Gaetan (2002). Dynamic generalized linear models with applications to environmental epidemiology. *Journal of the Royal Statistical Society Series C 51*, 453–468.

Ciocco, A. and D. Thompson (1961). A Follow-Up of Donora Ten Years After: Methodology And Findings. *American Journal of Public Health 51*, 155–164.

Clean Air Act (1990). U.S. Environmental Protection Agency.

Clean Air Act (1993). HM Stationary Office.

Cleveland, W. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association 74*, 829–836.

Conceicao, G., S. Miraglia, H. Kishi, P. Saldiva, and J. Singer (2001). Air Pollution and Child Mortality: A Time-Series Study in Sao Paula, Brazil. *Environmental Health Perspectives 109*, 347–350.

Consul, P. and G. Jain (1973). A Generalization of the Poisson Distribution. *Technometrics 15*, 791–799.

Cox, D. and H. Miller (1968). *The Theory of Stochastic Processes.* Chapman and Hall/CRC, London.

Cressie, N. and H. Huang (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association 94*, 1330–1340.

Dab, W., S. Medina, P. Quenel, Y. Le Moullec, A. Le Tertre, B. Thelot, C. Monteil, P. Lameloise, P. Pirard, R. Momas, R. Ferry, and B. Festy (1996). Short-term respiratory health effects of ambient air pollution: results of the APHEA project in Paris. *Journal of Epidemiology and Community Health 50*, S42–S46.

Daniels, M., F. Dominici, S. Zeger, and J. Samet (2004). The National Morbidity, Mortality, and Air Pollution Study Part III: Concentration-Response Curves and Thresholds for the 20 Largest US Cities. *HEI Project 96-97*, 1–21.

Davis, R., W. Dunsmuir, and S. Streett (2003). Observation-driven models for Poisson counts. *Biometrika 90*, 777–790.

Davis, R., W. Dunsmuir, and Y. Wang (1999). *Asymptotics, Nonparametrics and Time Series.* Marcel Dekker, New York.

Davis, R., W. Dunsmuir, and Y. Wang (2000). On autocorrelation in a Poisson regression model. *Biometrika 87*, 491–505.

Dellaportas, P. and A. Smith (1993). Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling. *Journal of the Royal Statistical Society Series C 42*, 443–459.

Department of the Environment, Transport and the Regions (1998). Review and assessment: monitoring air quality. *LAQM.TGI*.

Diggle, P., K. Liang, and S. Zeger (1994). *Longitudinal Data Analysis* (1st ed.). Oxford Science Publications, Oxford.

Dobson, A. (1990). *An Introduction to Generalized Linear Models* (1st ed.). Chapman and Hall/CRC, London.

Dockery, D. and C. Pope (1994). Acute Respiratory Effects Of Particulate Air Pollution. *Annual Reviews of Public Health 15*, 107–132.

Dockery, D., C. Pope, X. Xu, J. Spengler, J. Ware, M. Fay, B. Ferris, and F. Speizer (1993). An Association Between Air Pollution And Mortality In Six U.S. Cities. *The New England Journal of Medicine 329*, 1753–1759.

Dominici, F., M. Daniels, S. Zeger, and J. Samet (2002). Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships. *Journal of the American Statistical Association 97*, 100–111.

Dominici, F., J. Samet, and S. Zeger (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. *Journal of the Royal Statistical Society Series A 163*, 263–302.

Dominici, F., L. Sheppard, and M. Clyde (2003). Health Effects of Air Pollution: A Statistical Review. *International Statistical Review 71*, 243–276.

Dominici, F. and S. Zeger (2000). A measurement error model for time series studies of air pollution and mortality. *Biostatistics 1*, 157–175.

Duddek, C., N. Le, J. Zidek, and R. Burnett (1995). Multivariate imputation in cross-sectional analysis of health effects associated with air pollution. *Environmental and Ecological Statistics 2*, 191–212.

Efron, B. (1986). Double Exponential Families and Their Use in Generalized Linear Regression. *Journal of the American Statistical Association 81*, 709–721.

Eilers, P. and B. Marx (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science 11*, 89–1218.

Fahrmeir, L. (1992). Posterior Mode Estimation by Extended Kalman Filtering for Multivariate Dynamic Generalized Linear Models. *Journal of the American Statistical Association 87*, 501–509.

Fahrmeir, L., W. Hennevogl, and K. Klemme (1992). *Advances in GLIM and Statistical Modelling*. Springer, New York.

Fahrmeir, L. and H. Kaufmann (1991). On Kalman Filtering, Posterior Mode Estimation and Fisher Scoring in Dynamic Exponential Family Regression. *Metrika 38*, 37–60.

Fahrmeir, L. and S. Lang (2001). Bayesian inference for generalized additive mixed models based on Markov ranodm field priors. *Journal of the Royal Statistical Society Series C 50*, 201–220.

Fahrmeir, L. and G. Tutz (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models* (2nd ed.). Springer, New York.

Fahrmeir, L. and S. Wagenpfeil (1997). Penalized likelihood estimation and iterative Kalman smoothing for non-Gaussian dynamic regression models. *Computational Statistics and Data Analysis 24*, 295–320.

Firket, J. (1936). Fog Along The Meuse Valley. *Transactions of the Faraday Society 32*, 1192–1197.

Fruhwirth-Schnatter, S. (1994). Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering. *Statistics and Computing 4*, 259–269.

Fuentes, M. (2002). Interpolation of nonstationary air pollution processes: a spatial spectral approach. *Statistical Modelling 2*, 281–298.

Fuentes, M. and A. Raftery (2005). Model Evaluation and Spatial Interpolation by Bayesian Combination of Observations with Outputs from Numerical Models. *Biometrics 61*, 36–45.

Fuller, W. (1987). *Measurement Error Models* (1st ed.). John Wiley and Sons, Winchester.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika 85*, 215–227.

Gamerman, D. and H. Lopes (2006). *Markov chain Monte Carlo: Stochastic Simulation for Bayesian inference* (2nd ed.). Chapman and Hall/CRC, London.

Gelfand, A., L. Zhu, and B. Carlin (2001). On the change of support problem for spatio-temporal data. *Biostatistics 2*, 31–45.

Gelman, A. (2006). Prior distirbutions for variance parameters in hierarchical models. *Bayesian Analysis 1*, 515–533.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (2003). *Bayesian Data Analysis* (2nd ed.). Chapman and Hall/CRC, London.

Gelman, A., X. Meng, and H. Stern (1996). Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies. *Statistica Sinica 6*, 733–807.

Gelman, A. and D. Rubin (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science 7*, 457–472.

Geyer, C. and E. Thompson (1992). Constrained Monte Carlo Maximum Likelihood for Dependent Data. *Journal of the Royal Statistical Society Series B 54*, 657–699.

Gilks, W., S. Richardson, and S. D (1996). *Markov Chain Monte Carlo In Practice* (1st ed.). Chapman and Hall/CRC, London.

Gilks, W. and P. Wild (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Journal of the Royal Statistical Society Series C 41*, 337–348.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association 97*, 590–600.

Goldberg, M., R. Burnett, J. Bailar, J. Brook, Y. Bonvalot, R. Tamblyn, R. Singh, and M. Valois (2001). The Association between Daily Mortality and Ambient Air Particule Pollution in Montreal, Quebec. *Environmental Research Section A 86*, 12–25.

Gotway, C. and W. Stroup (1997). A generalised linear model approach to spatial data analysis adn prediction. *Journal of Agricultural, Biological and Environmental Statistics 2*, 157–178.

Greenland, S. and H. Morgenstern (1989). Ecological Bias, Confounding, and Effect Modification. *International Journal of Epidemiology 18*, 269–274.

Griffin, M. and K. Neuzil (2002). The Global Implications of Influenza in Hong Kong. *The New England Journal Of Medicine 347*, 2159–2162.

Gwynn, R., R. Burnett, and G. Thurston (2000). A Time-Series Analysis of Acidic Particulate Matter and Daily Mortality and Morbidity in the Buffalo, New York Region. *Environmental Health Perspectives 108*, 125–133.

Hastie, T. and R. Tibshirani (1986). Generalized Additive Models. *Statistical Science 1*, 297–318.

Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models* (1st ed.). Chapman and Hall/CRC, London.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient Models. *Journal of the Royal Statistical Society Series B 55*, 757–796.

Hastie, T. and R. Tibshirani (2000). Bayesian Backfitting. *Statistical Science 15*, 196–223.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

Hinkley, D., N. Reid, and E. Snell (1991). *Statistical Theory and Modelling* (1st ed.). Chapman and Hall/CRC, London.

Hoek, G., B. Brunekreef, S. Goldbohm, P. Fischer, and P. van-den Brandt (2002). Associations between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *Lancet 360*, 1203–1209.

Holloman, C., S. Bortnick, M. Morara, W. Strauss, and C. Calder (2004). A Bayesian Hierarchical Approach for Relating $PM_{2.5}$ Exposure to Cardiovascular Mortality in North Carolina. *Environmental Health Perspectives 112*, 1282–1288.

Hong, Y., J. Leem, E. Ha, and D. Christiani (1999). PM_{10} Exposure, Gaseous Pollutants, and Daily Mortality in Inchon, South Korea. *Environmental Health Perspectives 107*, 873–878.

Kalman, R. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering Series D 82*, 35–45.

Kass, R. and A. Raftery (1995). Bayes Factors. *Journal of the American Statistical Association 90*, 773–795.

Katsouyanni, K., J. Schwartz, C. Spix, G. Touloumi, D. Zmirou, A. Zanobetti, B. Wojtyniak, J. Vonk, A. Tobias, A. Ponka, S. Medina, L. Bacharove, and H. Anderson (1996). Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol. *Journal of Epidemiology and Community Health 50*, S12–S18.

Katsouyanni, K., G. Touloumi, E. Samoli, A. Gryparis, A. Tertre, Y. Monopolis, G. Rossi, D. Zmirou, F. Ballester, A. Boumghar, H. Anderson, B. Wojtyniak, A. Paldy, R. Braunstein, J. Pekkanen, C. Schindler, and J. Schwartz (2001). Confounding and Effect Modification in the Short-Term Effects of Ambient Particles on Total Mortality: Results from 29 European Cities within the APHEA2 Project. *Epidemiology 12*, 521–531.

Kelsall, J. and S. Zeger (1999). Frequency domain log-linear models; air pollution and mortality. *Journal of the Royal Statistical Society Series C 48*, 331–344.

Kitagawa, G. (1987). Non-Gaussian State-Space Modelling of Nonstationary Time Series. *Journal of the American Statistical Association 82*, 1032–1041.

Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State-Space Models. *Journal of Computational and Graphical Statistics 5*, 1–25.

Knorr-Held, L. (1999). Conditional Prior Proposals in Dynamic Models. *Scandinavian Journal of Statistics 26*, 129–144.

Laden, F., L. Neas, D. Dockery, and J. Schwartz (2000). Association of Fine Particulate Matter from Different Sources with Daily Mortality in Six U.S. Cities. *Environmental Health Perspectives 108*, 941–947.

Lang, S. and A. Brezger (2004). Bayesian P-Splines. *Journal of Computational and Graphical Statistics 13*, 183–212.

Lange, K. (1998). *Numerical Analysis for Statisticians* (3rd ed.). Springer, New York.

Lee, J., H. Kim, Y. Hong, H. Kwon, J. Schwartz, and D. Christiani (2000). Air Pollution and Daily Mortality in Seven Major Cities of Korea 1991-1997. *Environmental Research Section A 84*, 247–254.

Liang, K. and S. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Lin, C., M. Martins, S. Farhat, C. Pope, G. Conceicao, V. Anastacio, M. Hatanaka, W. Andrade, W. Hamaue, G. Bohm, and P. Saldiva (1999). Air pollution and respiratory illness of children in Sau Paulo, Brazil. *Paediatric and Perinatal Epidemiology 13*, 475–488.

Lin, M., Y. Chen, R. Burnett, P. Villeneuve, and D. Krewski (2002). The Influence of Ambient Coarse Particulate Matter on Asthma Hospitalization in Children: Case-Crossover and Time-Series Analysis. *Environmental Health Perspectives 110*, 575–581.

Lin, X. and D. Zhang (1999). Inference in Generalised Additive Mixed Models by using Smoothing Splines. *Journal of the Royal Statistical Society Series B 61*, 381–400.

Lioy, P., J. Waldman, T. Buckley, J. Butler, and T. Pietarinen (1990). The personal, indoor, and outdoor concentrations of PM_{10} measured in an industrial community during the wintger. *Atmospheric Environment 24*, 57–66.

Little, R. and D. Rubin (2002). *Statistical Analysis with Missing Data* (2nd ed.). John Wiley and Sons, Winchester.

Liu, J., W. Wong, and A. Kong (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika 81*, 27–40.

Lumley, T. and L. Sheppard (2000). Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses. *Environmetrics 11*, 705–717.

Mar, F., G. Norris, J. Koenig, and T. Larson (2000). Associations between Air Pollution and Mortality in Phoenix, 1995-1997. *Environmental Health Perspectives 108*, 347–353.

Marx, B. and P. Eilers (1998). Direct generalized additive modelling with penalized likelihood. *Computational Statistics and Data Analysis 28*, 193–209.

Matérn, B. (1986). *Spatial Variation* (2nd ed.). Springer, New York.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models* (2nd ed.). Chapman and Hall/CRC, London.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953). Equations of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics 21*, 1087–1092.

Miller, A. (1990). *Subset Selection in Regression* (1st ed.). Chapman and Hall/CRC, London.

Ministry of Public Health (1954). Mortality and Morbidity During the London Smog of December 1952. *HM Stationary Office*.

Moolgavkar, S. (2000). Air Pollution and Daily Mortality in Three U.S. Counties. *Environmental Health Perspectives 108*, 777–784.

Moolgavkar, S., H. Luebeck, T. Hall, and E. Anderson (1995). Air Pollution and Daily Mortality in Philadelphia. *Epidemiology 6*, 476–484.

Murray, C. and C. Nelson (2000). State-Space Modelling of the Relationship between Air Quality and Mortality. *Journal of the Air and Waste Management Association 50*, 1075–1080.

National Ambient Air Quality Standards (1990). US Environmental Protection Agency.

Neas, L., J. Schwartz, and D. Dockery (1999). A Case-Crossover Analysis of Air Pollution and Mortality in Philadelphia. *Environmental Health Perspectives 107*, 629–631.

Nelder, J. and R. Wedderburn (1972). Generalized Linear Models. *Journal of the Royal Statistical Society Series A 135*, 370–384.

Ott, W. (1990). A Physical Explanation of the Lognormality of Pollutant Concentrations. *Journal of the Air and Waste Management Association 40*, 1378–1383.

Ozkaynak, H., J. Xue, J. Spengler, L. Wallace, E. Pellizzari, and P. Jenkins (1996). Personal exposure to airborne particles and metals: Results from the particle team study in riverside, California. *Journal of Exposure Analysis and Environmental Epidemiology. 6*, 57–78.

Peng, R., F. Dominici, and T. Louis (2006). Model choice in time series of air pollution and mortality. *Journal of the Royal Statistical Society Series A 169*, 179–203.

Peng, R., F. Dominici, R. Pastor-Barriuso, S. Zeger, and J. Samet (2005). Seasonal Analyses of Air Pollution and Mortality in 100 U.S. Cities. *American Journal of Epidemiology 161*, 585–594.

Peng, R. D. and L. J. Welty (2004). The NMMAPSdata package. *R News 4*(2), 10–14.

Peters, A., J. Skorkovsky, F. Kotesovec, J. Brynda, C. Spix, E. Wichmann, and J. Heinrich (2000). Associations between Mortality and Air Pollution in Central Europe. *Environmental Health Perspectives 108*, 283–287.

Pettit, L. (1986). Diagnostics in Bayesian model choice. *The Statistician 35*, 183–190.

Plummer, M. and D. Clayton (1996). Estimation of Population Exposure in Ecological Studies. *Journal of the Royal Statistical Society Series B 58*, 113–126.

Pope, C., R. Burnett, M. Thun, E. Callee, D. Krewski, and K. Thurston (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association 287*, 1132–1141.

Pope, C. and D. Dockery (2006). Health Effects of Fine Particulate Air Pollution: Lines that Connect. *Journal of the Air and Waste Management Association 56*, 709–742.

Pope, C., M. Thun, M. Namboodiri, D. Dockery, J. Evans, F. Speizer, and C. Heath (1995). Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *American Journal of Respiratory and Critical Care Medicine 151*, 669–674.

R 2.1.0, a Language and Environment (2005). The R Development Core Team.

Ramsay, T., R. Burnett, and D. Krewski (2003). The Effect of Concurvity in Generalized Additive Models Linking Mortality to Ambient Particulate Matter. *Epidemiology 14*, 18–23.

Richardon, S., I. Stucker, and D. Hemon (1987). Comparison of Relative Risks Obtained in Ecological and Individual Studies: Some Methodological Considerations. *International Journal of Epidemiology 16*, 111–120.

Richardson, S. and W. Gilks (1993). Conditional Independence Models for Epidemiological Studies with Covariate Measurement Error. *Statistics in Medicine 12*, 1703–1722.

Richardson, S., L. Leblond, and P. Green (2002). Mixture models in measurement error problems, with reference to epidemiological studies. *Journal of the Royal Statistical Society Series A 165*, 549–566.

Ripley, B. (1987). *Stochastic Simulation* (1st ed.). John Wiley and Sons, Winchester.

Roberts, S. (2004). Biologically Plausible Particulate Air Pollution Mortality Concentration-Response Functions. *Environmental Health Perspectives 112*, 309–313.

Rubin, D. (1984). Bayesianly Justifiable And Relevant Frequency Calculations For The Applied Statistician. *Annals of Statistics 12*, 1151–1172.

Ruppert, D., M. Wand, and R. Carroll (2003). *Semiparametric Regression* (1st ed.). Cambridge University Press, Cambrdige.

Sahu, S., A. Gelfand, and D. Holland (2006). Spatio-temporal modelling of fine particulate matter. *Journal of Agricultural, Biological and Environmental Statistics 11*, 61–86.

Sahu, S. and K. Mardia (2005a). A Bayesian Kriged-Kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society Series C 54*, 223–244.

Sahu, S. and K. Mardia (2005b). Recent Trends in Modelling Spatio-Temporal data. *Proceedings of the special meeting on Statistics and Environment organized by the Società Italiana di Statistica held in Università Di Messina*, 69–83.

Samet, J., F. Dominici, F. Curriero, I. Coursac, and S. Zeger (2000). Fine Particulate Air Pollution and Mortality in 20 U.S. Cities 1987-1994. *The New England Journal of Medicine 343*, 1742–1749.

Samet, J., S. Zeger, F. Dominici, F. Curriero, I. Coursac, D. Dockery, J. Schwartz, and A. Zanobetti (2000). The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity and Mortality from Air Pollution in the United States. *HEI Project 96-97*, 5–47.

Schabenberger, O. and C. Gotway (2005). *Statistical Methods for Spatial Data Analysis.* Chapman and Hall/CRC, London.

Schafer, D. (2001). Semiparametric Maximum Likelihood for Measurement Error Model Regression. *Biometrics 57*, 53–61.

Schwartz, J. (1991). Particulate Air Pollution and Daily Mortality in Detroit. *Environmental Research 56*, 204–213.

Schwartz, J. (1993). Air Pollution and Daily Mortality in Birmingham, Alabama. *American Journal of Epidemiology 137*, 1136–1147.

Schwartz, J. (1994a). Air Pollution and Daily Mortality: A Review and Meta Analysis. *Environmental Research 64*, 36–52.

Schwartz, J. (1994b). Nonparametric smoothing in the analysis of air pollution and respiratory illness. *The Canadian Journal of Statistics 22*, 471–487.

Schwartz, J. (2000). The Distributed Lag between Air Pollution and Daily Deaths. *Epidemiology 11*, 320–326.

Schwartz, J. (2001). Is There Harvesting in the Association of Airborne Particles with Daily Deaths and Hospital Admissions. *Epidemiology 12*, 55–61.

Schwartz, J., F. Ballester, M. Saez, P. Perez-Hoyos, J. Bellido, K. Cambra, F. Arribas, A. Canada, M. Jose Perez-Boillos, and J. Sunyer (2001). The Concentration-Response Relation between Air Pollution and Daily Deaths. *Environmental Health Perpespectives 109*, 1001–1006.

Schwartz, J. and A. Marcus (1990). Mortality and Air Pollution in London: A Time Series Analysis. *American Journal of Epidemiology 131*, 185–194.

Schwartz, J., C. Spix, G. Touloumi, L. Bacharova, T. Barumamdzadeh, A. Le Tertre, T. Pierkarksi, A. Ponce de Leon, A. Ponka, G. Rossi, M. Saez, and J. Schouten (1996). Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *Journal of Epidemiology and Community Health 50*, S3–S11.

Schwartz, J. and A. Zanobetti (2000). Using Meta-Smoothing to Estimate Dose-Response Trends across Multiple Studies, with Application to Air Pollution and Daily Death. *Epidemiology 11*, 666–672.

Shaddick, G., L. Choo, and S. Walker (2007). Modelling correlated count data with covariates. *Journal of Statistical Computation and Simulation (to appear)*.

Shaddick, G. and J. Wakefield (2002). Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society Series C 51*, 351–372.

Shephard, N. and M. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–667.

Sheppard, L. and D. Damian (2000). Estimating short-term PM effects accounting for surrogate exposure measurements from ambient monitors. *Environmetrics 11*, 675–687.

Silvey, S. (1975). *Statistical Inference*. Chapman and Hall/CRC, London.

Smith, A. and G. Roberts (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society Series B 55*, 3–23.

Spiegelalter, D., N. Best, B. Carlin, and A. Van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society series B 64*, 583–639.

Spix, C., J. Heinrich, D. Dockery, J. Schwartz, G. Volksch, K. Schwinkowski, C. Collen, and E. Wichmann (1993). Air Pollution and Daily Mortality in Erfurt, East Germany, 1980-1989. *Environmental Health Perspectives 101*, 518–526.

Tonellato, S. (2001). A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations. *Journal of the Royal Statistical Society Series C 50*, 187–200.

UK Air Quality Strategy (2000). Department of the Environment, Transport and the regions.

Vedal, S., M. Brauer, R. White, and J. Petkau (2003). Air Pollution and Daily Mortality in a City with Low Levels of Pollution. *Environmental Health Perpespectives 111*, 45–51.

Verhoeff, A., G. Hoek, J. Schwartz, and J. van Wijnen (1996). Air Pollution and Daily Mortality in Amsterdam. *Epidemiology 7*, 225–230.

Wakefield, J. and R. Salway (2001). A statistical framework for ecological and aggregate studies. *Journal of the Royal Statistical Society Series A 164*, 119–137.

Wakefield, J. and G. Shaddick (2006). Health-exposure modelling and the ecological fallacy. *Biostatistics 7*, 438–455.

Wald, N. (2004). *The Epidemiological Approach* (4th ed.). Royal Society of Medicine Press, London.

Wang, P. and M. Puterman (1999). Markov Poisson regression models for discrete time series. Part 1: Methodology. *Journal of Applied Statistics 26*, 855–869.

West, M. and J. Harrison (1999). *Bayesian Forecasting and Dynamic Models* (2nd ed.). Springer, New York.

West, M., J. Harrison, and H. Migon (1985). Dynamic Generalized Linear Models and Bayesian Forecasting. *Journal of the American Statistical Association 80*, 73–83.

Wikle, C., M. Berliner, and N. Cressie (1998). Hierarchical Bayesian space-time mdoels. *Environmental and Ecological Statistics 5*, 117–154.

World Health Organisation (1975). Manual of the international Statistical Classification of Diseases, Injuries and Cause of Death, 9th revision.

Xu, X., B. Li, and H. Huang (1995). Air Pollution and Unscheduled Hospital Outpatient and Emergency Room Visits. *Environmental Health Perspectives 103*, 286–289.

Yu, O., L. Sheppard, T. Lumley, J. Koenig, and G. Shapiro (2000). Effects of Ambient Air Pollution on Symptoms of Asthma in Seattle-Area Children Enrolled in the CAMP Study. *Environmental Health Perspectives 108*, 1209–1214.

Zanobetti, A., J. Schwartz, E. Samoli, A. Gryparis, G. Touloumi, R. Atkinson, A. Tertre, J. Bobros, M. Celko, A. Goren, B. Forsberg, P. Michelozzi, D. Rabczenko, E. Ruiz, and K. Katsouyanni (2002). The Temporal Pattern of Mortality Responses to Air Pollution: A Multicity Assessment of Mortality Displacement. *Epidemiology 13*, 87–93.

Zanobetti, A., M. Wand, J. Schwartz, and L. Ryan (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics 1*, 279–292.

Zeger, S. (1988). A regression model for time series of counts. *Biometrika 75*, 621–629.

Zeger, S., F. Dominici, and J. Samet (1999). Harvesting-Resistant Estimates of Pollution Effects on Mortality. *Epidemiology 10*, 171–175.

Zeger, S. and B. Qaqish (1988). Markov Regression Models for Time Series: A Quasi-Likelihood Approach. *Biometrics 44*, 1019–1031.

Zeger, S., D. Thomas, F. Dominici, J. Samet, J. Schwartz, D. Dockery, and A. Cohen (2000). Exposure Measurement Error in Time-Series Studies of Air Pollution: Concepts and Consequences. *Environmental Health Perspectives 108*, 419–426.

Zhu, L., B. Carlin, and A. Gelfand (2003). Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics 14*, 537–557.

Zidek, J., G. Shaddick, R. White, J. Meloche, and C. Chatfield (2005). Using a probabilistic model (pcnem) to estimate personal exposure to air pollution. *Environmetrics 16*, 481–493.

Zidek, J., L. Sun, N. Le, and H. Ozkaynak (2002). Contending with space-time interaction in the spatial prediction of pollution: Vancouver's hourly ambient PM_{10} field. *Environmetrics 13*, 595–613.

Zmirou, D., J. Schwartz, M. Saez, A. Zanobetti, B. Wojtyniak, G. Touloumi, C. Spix, A. Ponce de Leon, Y. Le Moullec, L. Bacharova, J. Schouten, A. Ponka, and K. Katsouyanni (1998). Time-Series Analysis of Air Pollution and Cause Specific Mortality. *Epidemiology 9*, 495–503.